# PART I
# Getting Started

CHAPTER 1

*Introduction to Managerial Economics*

CHAPTER 2

*Basic Training*

INTEGRATING CASE STUDY

*Study I: Olsen's Preowned CD Players*

CHAPTER 1

# Introduction to Managerial Economics

## PREVIEW

For most purposes, economics can be divided into two broad categories: microeconomics and macroeconomics. *Macroeconomics* is the study of the economic system as a whole. It includes techniques for analyzing changes in total output, total employment, the consumer price index, the unemployment rate, and exports and imports. Macroeconomics addresses questions about the effect of changes in investment, government spending, and tax policy on exports, output, employment, and prices. Only aggregate levels of these variables are considered. But concealed in the aggregate data are countless changes in the output levels of individual firms, the consumption decisions of individual consumers, and the prices of particular goods and services.

Although macroeconomic issues and policies command much of the attention in newspapers and on television, the microdimensions of the economy are also important and are often more relevant to the day-to-day problems facing the manager. *Microeconomics* focuses on the behavior of the individual actors on the economic stage, that is, firms and individuals and their interaction in markets.

*Managerial economics* should be thought of as applied microeconomics. It is an application of the part of microeconomics that focuses on the topics that are of greatest interest and importance to managers. These topics include demand, production, cost, pricing, market structure, and government regulation. A strong grasp of the principles that govern the economic behavior of firms and individuals is an important managerial talent. The rational application of these principles should result in better managerial decisions, higher profits, and an increase in the value of the firm.

In general, managerial economics can be used by the goal-oriented manager in two ways. First, given an existing economic environment, the principles of managerial economics provide a framework for evaluating whether resources are being allocated efficiently within a firm. For example, economics can help the manager determine if profit could be increased by reallocating labor from a marketing activity to the production line. Second, these principles help managers respond to various economic signals. For example, given an increase in the price of output or the development of a new lower-cost production technology, the appropriate managerial response would be to increase output. Alternatively, an increase in the price of one input, say labor, may be a signal to substitute other inputs, such as capital, for labor in the production process.

The tools developed in the following chapters will increase the effectiveness of decision making by expanding and sharpening the analytical framework used by managers to make decisions. Thus, a working knowledge of the principles of managerial economics can increase the value of both the firm and the manager.

This chapter sets the stage for the development of managerial economic skills. First, the interrelationships among consumers, firms, and resource owners in a market economy are outlined. Next, the nature and objective of the firm and the importance of profit as an incentive for firms to respond to consumer demands for output are discussed. Following that, the potential for owners and managers to have different objectives (i.e., the principal-agent problem) is discussed. Finally, the role of economics in decision making is considered.

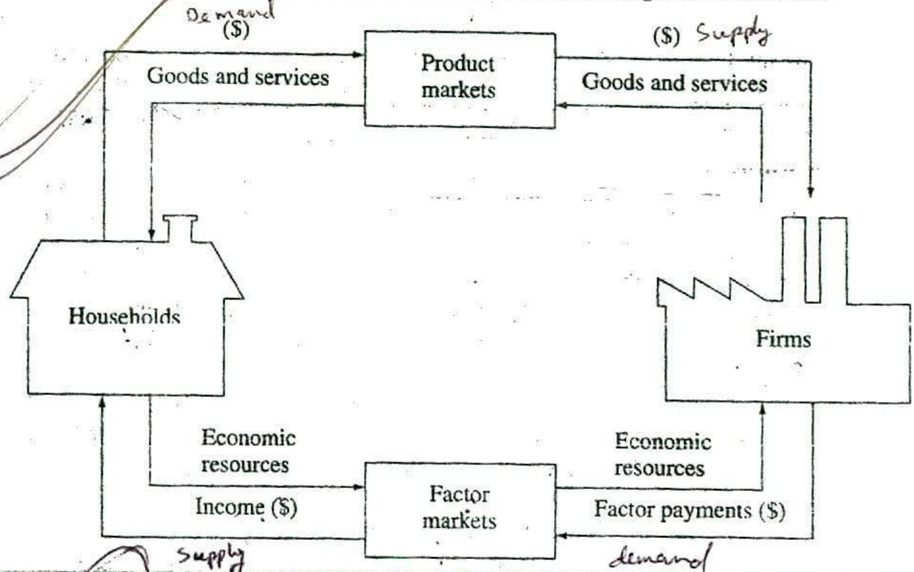



FIGURE 1.1 Circular Flow of Income, Output, Resources, and Factor Payments

## THE CIRCULAR FLOW OF ECONOMIC ACTIVITY

Individuals and firms are the fundamental participants in a market economy. Individuals own or control resources that have value to firms because they are necessary inputs in the production process. These resources are broadly classified as labor, capital, and natural resources. Of course, there are many types and grades of each resource. Labor specialties vary from street sweepers to brain surgeons; capital goods range from brooms to computers. Most people have labor resources to sell, and many own capital and/or natural resources that are rented, loaned, or sold to firms to be used as inputs in the production process. The money received by an individual from the sale of these resources is called a factor payment. This income to individuals then is used to satisfy their consumption demands for goods and services.

The interaction between individuals and firms occurs in two distinct arenas. First, there is a product market where goods and services are bought and sold. Second, there is a market for factors of production where labor, capital, and natural resources are traded. These interactions are depicted in Figure 1.1, which describes the circular flow of income, output, resources, and factor payments in a market economy.

In the product market shown in the top part of the figure, individuals demand goods and services in order to satisfy their consumption desires.[1] They make these demands known by bidding in the product market for these goods and services. Firms, anxious to earn profits, respond to these demands by supplying goods and services to that market. The firm's production technology and input costs determine the supply conditions, while consumer preferences and income (i.e., the ability to pay) determine the demand conditions.

[1] It is necessary to distinguish between *want* and *demand*. Many individuals *want* goods and services but cannot afford to buy them. The term *demand* implies that the consumer has both the desire to have the goods and the ability to buy it.

The interaction of supply and demand determines the price and quantity sold. In the product market, purchasing power, usually in the form of money, flows from consumers to firms. At the same time, goods and services flow in the opposite direction—from firms to consumers.

The factor market is shown at the bottom of Figure 1.1. Here, the flows are the reverse of those in the product market. Individuals are the suppliers in the factor market. They supply labor services, capital, and natural resources to firms that demand them to produce goods and services. Firms indicate the strength of their desire for these inputs by bidding for them in the market. The flow of money is from firms to individuals, and factors of production flow from individuals to firms. The prices of these productive factors are set in this market.

Prices and profits serve as the signals for regulating the flows of money and resources through the factor markets and the flows of money and goods through the product market. For example, relatively high prices and profits in the personal computer industry in the 1980s signaled producers to increase production and send more units of output to the product market. To produce more computers, more labor and capital were required. Firms raised the prices they would pay for these resources in the factor market to signal resource owners that higher returns were now available. The result was rapid growth in the personal computer industry as resources were bid away from other industries. In the early 1990s, this market became very competitive and prices fell substantially. Although total unit sales increased, the profit on each computer was smaller, and some firms struggled to keep total profit at an acceptable level. Consumers benefited greatly as better and more powerful computers became available at lower and lower prices.

In the market economy depicted by this circular flow, individuals and firms are highly interdependent; each participant needs the others. For example, an individual's labor will have no value in the market unless there is a firm that is willing to pay for it. Alternatively, firms cannot justify production unless some consumers want to buy their products. As a result, all participants have an incentive to provide what others want. All participate willingly because they have something to gain by doing so. Firms earn profits, the consumption demands of individuals are satisfied, and resource owners receive wage, rent, and interest payments. If some do not benefit by buying and selling in these markets, they are not required to do so. Thus, one can be sure that no individual is made worse off by voluntary trade in these factor and product markets. Indeed, the gains that accrue to the individual participants form the essence of a market economy.

## THE NATURE OF THE FIRM

In order to earn profits, the firm organizes the factors of production to produce goods and services that will meet the demands of individual consumers and other firms. The concept of the firm plays a central role in the theory and practice of managerial economics. Thus, a significant part of this text is focused on production, cost, and the organization of firms in the marketplace. These topics form the basis for what is known in economics as the theory of the firm. An understanding of the reason for the existence of firms, their specific role in the economy, and their objective provides a background for that theory.

## The Rationale for the Firm[2]

In a free-market economy, the organization and interaction of producers (i.e., firms) and consumers is accomplished through the price system. There is no need for any central direction by government, nor is such central control or planning thought to be desirable. Within the firm, however, transactions and the organization of productive factors are generally accomplished by the central control of one or more managers. For example, workers subject themselves almost completely to management during the work period. Thus, there is an apparent dichotomy in the organization of production in a market economy. The price system guides the decentralized interaction among consumers and firms, whereas central planning and control tend to guide the interaction within firms. This raises the question, Why is the production system not completely guided by price signals? That is, why do firms exist in a market economy?

Essentially, firms exist as organizations because the total cost of producing any rate of output is lower than if the firm did not exist. There are several reasons why these costs are lower. First, there is a cost of using the price system to organize production. The cost of obtaining information on prices and the cost of negotiating and concluding separate contracts for each step of the production process would be very burdensome. Firms often hire labor for long periods of time under agreements that specify only that a wage rate per hour or day will be paid for the workers doing what they are asked. That is, one general contract covers what usually will be a large number of transactions between the owners and workers. The two parties do not have to negotiate a new contract every time the worker is given a new assignment. The saving of the transactions costs associated with such negotiations is advantageous to both parties, and thus both labor and management voluntarily seek out such arrangements.

A secondary explanation for the existence of firms is that some government interference in the marketplace applies to transactions among firms rather than within firms. For example, sales taxes usually apply only to transactions between one firm and another. In some states, a construction company may have to pay sales tax on cabinets purchased from an independent cabinetmaker. By hiring that person, this tax is avoided and the cost of producing output is reduced. By internalizing some transactions within the firm that would otherwise be subject to those interferences, production costs are reduced. Because this is a secondary factor, firms would exist in the absence of such interference, but it probably contributes to the existence of more and larger firms.

Given that production costs are reduced by organizing production factors into firms, why won't this process continue until there is just one large firm, such as a giant General Motors or Exxon that produces all goods and services for the entire economy? There are at least two reasons. First, the cost of organizing transactions within the firm tends to rise as the firm gets larger. Logic dictates that the firm will internalize the lower-cost transaction first, and then the higher-cost transactions. At some point, these internal transactions costs will equal the costs of transacting in the market. At that point, the firm will cease to grow. For example, all automobile producers in the world buy tires from companies that specialize in the production of rubber products. Surely,

---

[2]This section draws on the classic article by R. H. Coase, "The Nature of the Firm," *Economica* (November 1937): 386–405.

Ford and General Motors must have considered building plants to produce their own tires. It can be inferred that the cost of developing the new management skills required for such a different type of production, and the difficulty of managing an even larger and more complicated business, must have been greater than the cost of continuing to buy tires from Goodyear, Michelin, or other producers.

Another example is legal services. Usually, attorneys are not an integral part of the production process, but are needed periodically. It would be too costly for many firms to have full-time attorneys whose services would not be needed on a continual basis. So, rather than having a full-time lawyer employed by the firm, legal services are contracted on a when-needed basis. The cost of such an arrangement for most firms is lower. In contrast, large firms that have a continual need for legal services generally have an in-house legal staff, but even they make extensive use of outside law firms.

A second factor constraining firm size is the limitation of an entrepreneur's organizational skill. Resources within the company may not be efficiently allocated if the firm's size exceeds the manager's ability to control the operation. To overcome this problem, many large firms are organized into groups of divisions referred to as *profit centers*. The management of each of these seeks to maximize that division's profit. By having a number of smaller organizations, each being managed somewhat independently, the problem of limited ability to control the larger firm is at least partially overcome.

Both of these reasons for a limit on the size of the firm fall under the heading of what economists have termed *diminishing returns to management*. Stated another way, production costs per unit of output will tend to rise as firms grow larger, because of limited managerial ability. It should be noted that many large firms recognize the problem of excessive size and decentralize by establishing a number of separate divisions or profit centers that act as individual firms.

**Key Concepts**

- The interaction of individuals and firms in a market economy can be described as a circular flow of money, goods and services, and resources through product and factor markets.
- Firms exist because the costs of production are lower and returns to the owners of labor and capital are higher than if the firm did not exist.
- Limits are imposed on the size of firms because the cost of organizing transactions rises as the firm becomes larger and because managerial ability is limited.

## The Objective of the Firm

This book approaches microeconomics from the perspective of efficient management of a business or other organization. To be able to discuss efficient or optimal decision making requires that a goal or objective be established. That is, a management decision can only be evaluated against the goal that the firm is attempting to achieve. Traditionally, economists have assumed that the objective of the firm is to maximize profit. That is, it is assumed that managers consistently make decisions in order to maximize profit. But profit in which period? This year? The next five years? Often, managers are observed making decisions that reduce current year profits in an effort to increase profits in future years. Expenditures for research and development, new capital equipment, and major marketing programs are but a few examples of activities that reduce profits initially but will significantly increase profits in later years.[3]

As both current and future profits are important, it is assumed that the goal is to maximize the present or discounted value of all future profits.[4] Formally stated, the goal or objective function for the firm is to

$$\text{maximize: } PV(\pi) = \frac{\pi_1}{1+r} + \frac{\pi_2}{(1+r)^2} + \cdots + \frac{\pi_n}{(1+r)^n} \qquad (1\text{–}1)$$

where $\pi_t$ is profit in time period $t$, and $r$ is an appropriate discount rate used to reduce future profits to their present value. Using the Greek letter $\Sigma$ to indicate that each of the terms on the right-hand side of the equation have been added together, the objective function can be written as

$$\text{maximize: } PV(\pi) = \sum_{t=1}^{n} \frac{\pi_t}{(1+r)^t}$$

The present value of all future profits also can be interpreted as the value of the firm, that is, what a willing buyer would pay for the business. Thus, to maximize the discounted value of all future profits is equivalent to maximizing the value of the firm. The terms *profit maximization* and *value maximization* will be used interchangeably in the remainder of this book.

Most of the large firms in market economies are corporations where ownership is spread among literally thousands of individuals, each of whom owns shares of stock that represent that ownership. These owners elect a board of directors that, in turn, hires executives who will manage the firm. It is now common to hear these managers talk in terms of making decisions that will maximize shareholder value. This is simply another way of stating the goal of profit maximization.

## Maximizing versus Satisficing

Considerable controversy has developed about the "realism" of this assumed management goal. Advocates of alternative theories of the firm argue that the behavior of real-world managers is not always consistent with the profit-maximization goal. Other objectives are seen as being at least as important. Among the alternatives that have been identified are maximizing total revenue, maximizing employment tenure and departmental budget, maximizing executive salaries, achieving "satisfactory" profit levels, maximizing the manager's individual utility function, and maximizing market share subject to a satisfactory profit constraint. Indeed, an entire literature has developed on the subject of "satisficing" rather than "maximizing" management objectives.[5]

---

[3]Some managers who must report profit performance monthly or quarterly claim that the pressure for increased short-term profits may cause them to make decisions that increase these profits at the expense of long-term profit.

[4]Many students will have already studied the concept of present value. For those who have not, this principle is developed more fully in the appendix to this chapter.

[5]See H. A. Simon, "Theories of Decision Making in Economics," *American Economic Review* 49, no. 3 (1959):253–283; and R. M. Cyert and J. A. March, *A Behavioral Theory of the Firm* (Upper Saddle River, NJ: Prentice-Hall, 1963).

Some critics of the profit-maximization assumption argue that it is unrealistic because managers must function in an environment characterized by inadequate information and uncertainty about the outcome of any strategy that might be adopted. Therefore, as a practical matter, it is really impossible to maximize profit. Economist Fritz Machlup responded to this criticism by comparing managing a firm to driving a car.[6] When deciding whether to pass another car on a two-lane highway, a driver must consider a variety of conditions. A few of these include: the speed of the driver's car, the car that is to be passed, and that of any oncoming vehicles; the road conditions; the existence of any curves or intersections that may be upcoming; and lighting conditions. Obviously, to analyze this situation completely prior to passing the car ahead might require the assistance of a computer and several working days.

Clearly, however, the typical driver makes an intuitive evaluation of these conditions in a matter of a few seconds or less and makes a decision. Almost all of the time, it is the correct decision. Machlup contends that managers are seeking to maximize profit and simply react intuitively to a set of conditions characterized by incomplete information and uncertainty in the same manner as the automobile driver. Thus, while there may be additional information that would result in even greater profit, managers are striving to generate the greatest possible profit given the limited information available.

Although some managers may have other goals, most of the criticism leveled at the profit-maximization assumption may be irrelevant. Economics is less interested in how some managers really act than in understanding the economic environment in which managers must function and, more importantly, in developing a framework for predicting managerial responses to important changes in that environment.

In this context, Machlup argues that it is fruitless to worry about how real-world business managers really behave and what their goals really are. The essential question is: If we assume that firms attempt to maximize profit, will the principles of economics derived from that objective function explain the behavior of real-world firms? If the answer is yes, it does not matter if the assumed goal is not entirely realistic. For example, suppose that there are two firms, A and B, that are identical except in one respect: The manager of A works 16 hours a day, seven days a week, whereas the manager of B works only four days each week, and leaves the office shortly after lunch each day to play golf. Because of manager A's commitment to work, it is probable that profits for firm A are greater than at firm B, but this difference in commitment to work and its effect on profit is not the issue. The fundamental question is: Will the two managers respond the same way to changes in economic conditions? If the price of their product increases, do both use the profit-maximizing strategy of increasing the rate of output? The manager of A may take action first, because he is on the job most of the time, but the response should be the same for both firms.

Most economists agree that the principles of managerial microeconomics do indeed allow accurate predictions of managerial decisions, and that profit maximization provides a useful assumption in that context. Indeed, no general theory has yet been proven to predict more accurately than the models based on profit maximization. Thus, in this book, it will be assumed that the objective of the firm is to maximize profit or, equivalently, to maximize the value of the firm.

---

[6] F. Machlup, "Theories of the Firm: Marginalist, Behavioral, Managerial," *American Economic Review* 57, no. 1 (1967):1–33.

## The Principal-Agent Problem

In the modern corporation, the owners or stockholders (i.e., the principals) hire managers (i.e., agents) to conduct the day-to-day operation of the firm. These managers are paid a salary to represent the interest of the owners, ostensibly, to maximize the value of the firm. A board of directors is elected by the owners to meet regularly with the managers to oversee their activity and to try to ensure the managers are, in fact, acting in the best interest of the owners.

Because of the difficulty of monitoring the managers on a continual basis, it is possible that goals other than profit-maximization may be pursued. In addition to those mentioned earlier, the managers may seek to enhance their positions by spending corporate funds on fancy offices, excessive and expensive travel, club memberships, and so forth. In recent years, many corporations have taken action to align the interests of owners with the interests of the managers by tying a large share of managerial compensation to the financial performance of the firm.

For example, the manager may be given a basic salary plus potentially large bonuses for meeting such goals as attaining a specified return on capital, growth in earnings, and/or increase in the price of the firm's stock. With regard to the latter, the use of stock options awarded to top managers is a most effective way to insure that managers act in the interest of the shareholders. Typically, the arrangement provides that the manager is to receive an option to buy a specified number of shares of common stock at the current market price for a specified number of years. The only way the executives can benefit from such an arrangement is if the price of stock rises during the specified term. The option is exercised by buying the shares at the specified price, and the gain equals the increase in share price multiplied by the number of shares purchased. Sometimes the agreement specifies that the stock must be held for several years following purchase.

Essentially, this option arrangement makes the manager a *de facto* owner, even if the option has not been exercised. In almost every case of a report of unusually high executive compensation, the largest part of that compensation is associated with gains from stock options.

Consider the following example. Smith is hired as the president of a firm at an annual salary of $500,000 plus a five-year option to buy 100,000 shares of stock at the current market price of $50 per share. Assume that within five years the price of the stock has increased to $75 per share. Smith exercises the option by buying 100,000 shares for $5,000,000, which have a market value of $7,500,000. In the year the options are exercised, Smith has a gain (i.e., additional compensation) of $2,500,000. However, if the price of the stock had remained unchanged or had declined, this option would have no value and Smith would have received no additional compensation.[7]

[7] A study at the University of Washington (Narayanan, 1996) found that managers who were paid only a cash salary (no stock options) tended to make decisions that resulted in less investment that would increase future profits. In contrast, those whose contracts were heavily weighted in common stock grants and options tended to overinvest. It was shown that management contracts that provided for both cash and stock tended to produce efficient long-term investment rates. See M. P. Narayanan, "Form of Compensation and Managerial Decision Horizon," *Journal of Financial and Quantitative Analysis* 31, no. 4 (December 1966): 467–491.

### *Executive Pay in the United States*

If there were any questions about the potential rewards to the top executives at America's largest corporations, consider the levels of compensation of the 10 highest-paid executives in 1996:

| | *1996 Salary and Bonus* | *Long-term Compensation* | *Total Pay* |
|---|---|---|---|
| **Lawrence Coss** Green Tree Financial | $102,499,000 | none | $102,499,000 |
| **Andrew Grove** Intel | 3,003,000 | $94,587,000 | 97,590,000 |
| **Sanford Weill** Travelers Group | 6,330,000 | 87,828,000 | 94,158,000 |
| **Theodore Waitt** Gateway 2000 | 965,000 | 80,361,000 | 81,326,000 |
| **Anthony O-Reilly** H. J. Heinz | 2,736,000 | 61,500,000 | 64,236,000 |
| **Sterling Williams** Sterling Software | 1,448,000 | 56,801,000 | 58,249,000 |
| **John Reed** Citicorp | 3,467,000 | 40,143,000 | 43,610,000 |
| **Stephen Hilbert** Conseco | 13,962,000 | 23,450,000 | 37,412,000 |
| **Casey Cowell** U. S. Robotics | 3,430,000 | 30,522,000 | 33,952,000 |
| **James Moffett** Freeport-McMoran C&G | 6,956,000 | 26,776,000 | 33,732,000 |

Note that except for Lawrence Coss, by far, the largest part of compensation comes in the form of long-term compensation, usually stock options.

Bill Gates, the CEO at Microsoft, was paid "only" about $350,000. Lest one worry about his financial condition, the value of his 300 million shares of stock increased by $11.4 billion during the year!

## Constrained Decision Making

The essence of the science of economics is determining optimum behavior where that behavior is subject to constraints. Profit maximization is constrained by the limited information available to the manager. In general, constraints on managerial decisions involve legal, moral, contractual, financial, and technological considerations. Legal constraints include the array of federal, state, and local laws that must be obeyed by all citizens, both individual and corporate. Areas where managers seem to be having some legal difficulty include environmental laws, especially those relating to pollution and the disposal of hazardous wastes, and employment law, including wrongful termination and sexual harassment matters. Moral constraints apply to actions that are not illegal but are sufficiently inconsistent with generally accepted standards of behavior to be considered improper. Contractual constraints bind the firm because of some prior agreement such as a long-term lease on a building, or a contract with a labor union that represents the firm's employees.

Finally, there are financial and technological constraints. An example of a financial constraint occurs when a department of a firm is assigned a budget for the next year and managers are given orders such as "maximize production subject to this budgeted amount." Technological constraints set physical limits on the amount of output per unit of time that can be generated by particular machines or workers. As will be shown in subsequent chapters, making optimizing decisions under financial or technological constraints is a fundamental part of managerial economics.

## *Case Study*

### Adam Smith and the Invisible Hand

The eighteenth-century philosopher Adam Smith is regarded as the father of modern economics. He understood clearly how a decentralized market economy would function with little, if any, outside regulation required. In particular, he saw that by seeking to maximize their self-interest, the individual participants in this economy would achieve socially desirable results. In his book *An Inquiry into the Nature and Causes of the Wealth of Nations** (originally published in 1776), Smith writes

> It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own self-interest. We address ourselves, not to their humanity but to their self-love, never talk to them of our own necessities, but of their advantages.

The coordinating mechanism in the economy, the price system, was described by Smith as an "invisible hand" that guided private decisions in socially beneficial ways:

> Every individual endeavors to employ his capital so that its produce may be of greatest value. He generally neither intends to promote the public interest, nor knows how much he is promoting it. He intends only his own security, only his own gain. And he is in this led by an invisible hand to promote an end which was no part of his intention. By pursuing his own interest he frequently promotes that of society more effectually than when he really intends to promote it.

In addition to this rationale for a market system unfettered by outside (e.g., government) influence or control, Smith's example of the efficiencies associated with the mass production of pins provides insight into the role of the firm in market economy and the advantages of mass production.

---

*New York: Random House/Modern Library, 1985.

> To take an example . . . the trade of the pinmaker; a workman not educated to this business . . . nor acquainted with the use of the machinery employed in it . . . could scarce, perhaps, with his utmost industry, make one pin in a day, and certainly could not make twenty. But in the way in which this business is now carried on not only the whole work is a peculiar trade, but it is divided into a number of branches, of which the greater part are likewise peculiar trades. One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; . . . and the important business of making a pin is, in this manner, divided into about eighteen distinct operations. . . . I have seen a small manufactory of this kind where ten men only were employed making upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in a day.

Even in 1776, Adam Smith could see the advantages of an economic system based on the self-interest of the individual participants and the development of firms to organize labor and capital for the purpose of low-cost mass production. ■

**Key Concepts**

- It is assumed that the objective of the firm is to maximize the present value of all future profits, subject to various legal, moral, contractual, financial, and technological constraints.
- The principles of managerial economics allow accurate predictions of decision making in business and other organizations.
- The principal-agent problem refers to the possibility that owners and their managers may have different objectives. These interests can be aligned through the use of managerial compensation arrangements that tie individual compensation to the overall performance of the firm.

## THE CONCEPT OF ECONOMIC PROFIT

Having determined that the goal of the firm is profit maximization, it is necessary to define the term *profit*. The conventional notion of profit is relatively straightforward: Profit is defined as revenues minus costs. However the definition of cost is quite different for the economist than for the accountant. Consider an individual who has an MBA degree and is considering investing $200,000 in a retail store that she would manage. The projected income statement for the year as prepared by an accountant is as shown here.

| | | |
|---|---|---|
| Sales | | $90,000 |
| Less: Cost of goods sold | | 40,000 |
| Gross profit | | $50,000 |
| Less: Advertising | $10,000 | |
| Depreciation | 10,000 | |
| Utilities | 3,000 | |
| Property tax | 2,000 | |
| Miscellaneous expenses | 5,000 | 30,000 |
| Net accounting profit | | $20,000 |

This accounting or business profit is what is reported in publications such as the *Wall Street Journal* and in the quarterly and annual financial reports of businesses. It is a meaningful concept as far as it goes—it just does not go far enough. Furthermore, the use of this concept may result in making the wrong decision.

The economist recognizes other costs, defined as *implicit costs*. These costs are not reflected in cash outlays by the firm, but are the costs associated with foregone opportunities. Such implicit costs are not included in the accounting statements, but must be included in any rational decision-making framework. There are two major implicit costs in the preceding example. First, the owner has $200,000 invested in the business. Suppose the best alternative use for this money is a bank account paying a 5 percent interest rate. Therefore, this investment would return $10,000 annually. Thus, $10,000 should be considered as the implicit or *opportunity cost* of having the $200,000 invested in the retail store.

The second implicit cost includes the manager's time and talent. The annual wage return on an MBA degree from a reasonably good business school may be $60,000 per year. This is the implicit cost of managing this business rather than working for someone else. Thus the income statement should be amended in the following way in order to determine economic profit:

| | | |
|---|---|---|
| Sales | | $90,000 |
| Less: Cost of goods sold | | 40,000 |
| Gross profit | | $50,000 |
| Less: Explicit costs: | | |
| Advertising | $10,000 | |
| Depreciation | 10,000 | |
| Utilities | 3,000 | |
| Property tax | 2,000 | |
| Miscellaneous expenses | 5,000 | 30,000 |
| Accounting profit (i.e., profit before implicit costs) | | $20,000 |
| Less: Implicit costs: | | |
| Return on $200,000 of invested capital | 10,000 | |
| Foregone wages | 60,000 | 70,000 |
| Net "economic profit" | | $-50,000 |

From this broader perspective, the business is projected to lose $50,000 in the first year. The $20,000 accounting profit disappears when all "relevant" costs are included. Obviously, with the financial information reported in this way, an entirely different decision might be made on whether to start this business. Another way of looking at the problem is to assume that $200,000 had to be borrowed at 5 percent interest and an MBA graduate hired at $60,000 per year to run the store. In this case, the implicit costs become explicit, and the accounting profit is the same as the economic profit (i.e., $-50,000) because all costs, both explicit and implicit, have been considered.

Most decision makers are aware of this concept. The carpenter who works as an independent contractor rather than as an employee of another firm knows that his opportunity cost is the market wage rate for carpenters. The contractor's lament, "I lost money on that job" may mean that his accounting profit was $100 per day when he could have made $150 per day working for someone else.

The important point of this discussion is that an entirely different signal may be given to management when the concept of economic profit is used. Sometimes the operation of economically unprofitable businesses is continued because of a failure to understand and properly include implicit costs. Rational decision making requires that all relevant costs, both explicit and implicit, be recognized. The concept of economic profit accounts for all costs and therefore is a more useful management tool than the more normally defined concept of accounting profit. In the following chapters, the economic definition of profit will be used, and the term *cost* is defined to include all relevant costs, both explicit and implicit.

## Example Opportunity Costs

Sharon Smith is a full-time homemaker and is also an excellent seamstress. She has material for which she paid $5 per yard several years ago. The material has increased in value during that time and could be sold back to the local fabric shop for $15 per yard. Sharon is considering using that material to make dresses, which she would sell to her friends and neighbors. She estimates that each dress would require four yards of material and four hours of her time, which she values at $10 per hour. If the dresses could be sold for $90 each, could Sharon earn a positive economic profit by making and selling the dresses?

**Solution** The key to this decision is appropriately accounting for both Sharon's time ($10 per hour) and the true opportunity cost of the material, $15 per yard (the amount she could receive by selling it to the fabric shop). The profit calculation per dress would be as follows:

| | |
|---|---|
| Revenue | $90 |
| Less: 4 hours of labor at $10/hr | 40 |
| 4 yards of material at $15/yd | 60 |
| Economic profit | $-10 |

Clearly, making the dresses is not going to be profitable.

If Sharon had not included the value of her time and had used the historic price of $5 per yard as the cost of the material, she would have estimated a "profit" of $70 per dress, that is

| | |
|---|---|
| Revenue | $90 |
| Less: 4 yards of material at $5/yd | 20 |
| "Profit" | $70 |

This is not an accurate measure of profit because it fails to account for the true opportunity cost of Sharon's time and the opportunity cost of a yard of material. She could sell both in the market and make more than she could by producing the dresses.

## PROFIT IN A MARKET SYSTEM

Except for those who espouse the view that charity is the only virtue, contributors to output expect to be paid. Each workday, many people willingly spend about one-half of their waking hours in employment for which they are paid wages. Those who own land and structures will let others use those resources as long as rent is paid for their use. Owners of financial resources make their assets available in return for the payment of interest.

Of the total income generated in the U.S. economy each year, about 73 percent is in the form of wages. Rent accounts for 2 percent and interest comprises about 10 percent. Profit, although representing only 15 percent of national income, plays two critically important roles in the functioning of the economic system. First, profit acts as a signal to producers to change the rate of output, or to enter or leave an industry. Second, profit is a reward that encourages entrepreneurs to organize factors of production and take risk.

High profits in an industry are usually a signal that buyers want more output from that industry. Those profits provide the incentive for existing firms to increase output and for new firms to enter the market. Conversely, low profits are a signal that less output is being demanded by consumers and/or that production methods are not efficient. When demand for output decreases, firms reduce production of that product, and resources are made available for the production of other goods and services for which demand has increased. If low profits are an indication of inefficiency in production, rather than inadequate demand, management is signaled to reorganize the production process or otherwise reduce costs.

It seems clear that profit is the objective of most managers and entrepreneurs. Although sometimes characterized by greed, avarice, misrepresentation, or outright lawbreaking, the desire for profit drives the market economy. This quest for profit is no more selfish than the quest for higher wages. Profit is a residual income after other participants in production have been paid. The sellers of labor inputs want their wages before the final goods are sold to consumers. Generally, they do not want to share the risks. Thus, it is not surprising that entrepreneurs will not commit time and resources to risky activities unless there is the prospect of earning a profit. One function of profit is to reward entrepreneurs for accepting the risk associated with their business decisions. Virtually all of these decisions carry the risk that money will be lost. Clearly, if there is a chance of a loss, there must be the chance for a gain in the form of profit; otherwise, such risks will not be taken.

Certain firms seem to be able to earn above-normal profits on a consistent basis. In some cases, they are continually innovative in developing new products, in reducing production costs on existing products, or in providing unusually good customer service. Although other firms may follow their lead and copy any new product or duplicate a production-cost advantage, over time the truly innovative firm will have developed even more new products and production technology, thus maintaining its above-normal profit performance. In such cases, profit plays a useful social function; it encourages firms to develop new products, to lower production costs, and to provide better service.

In other cases, firms may earn above-normal profit because they have monopoly power in a market. That is, other firms are prevented from offering the same product or service for sale, thus allowing that firm to maintain high prices and profits. For example, Resorts International, Inc., was the first firm to be granted a gambling license in Atlantic City, New Jersey, when that activity was legalized in 1978. For three years, the firm made large profits. As soon as licenses were given to other firms, competition increased and Resorts' profit fell more than 50 percent. In another case, Bausch & Lomb, a large manufacturer of precision lenses, was given permission by the U.S. Food and Drug Administration (FDA) to produce and sell soft contact lenses. For several years the firm had a monopoly because no other firm was able to obtain FDA permission. During that period, prices of these lenses were very high, as were profits for Bausch & Lomb. As soon as other firms were able to enter the market, both prices and profits declined.

In cases such as these, the above-normal profit is not socially useful. Indeed, a firm with monopoly power does not have as much incentive to hold down prices and production costs, to develop new products, or to provide good customer service as does a firm in a very competitive industry. However, in some industries, the lowest production cost can be achieved by having only one firm in the market. The generation and delivery of electric power is an example of such an industry. In most areas of the country, one firm is given a monopoly franchise but is subject to government regulation of price to ensure that "reasonable" prices are charged and that only a "fair" profit is earned. In this way, the advantage of low production cost is achieved without the above-normal profits usually associated with monopoly power.

Some critics of the capitalist system argue that profit is little more than a windfall gain that is randomly conferred on certain people. As such, they believe that it should be taxed away because it is "unearned." However, even if they were correct in claiming that profit is no more than a random windfall, the free enterprise system cannot work without profit. That is, no one will "play the game" (i.e., make investments in business) if there is no chance of winning a prize in the form of profit.

In a market characterized by many firms competing against one another, above-normal profits provide important signals, but are not likely to be maintained over long periods of time. That is, firms already in the market respond to higher profits by increasing output, and new firms will have an incentive to enter the market as well. The result will be an increased supply of the product, lower prices, and, ultimately, lower profits. The result in competitive markets is that profits provide important signals, but are somewhat transitory in nature.

## ase Study

### illiam Henry Gates III and the Microsoft Money Machine

Several years ago, when his fortune was a mere several hundred million dollars, a weekly magazine labeled Bill Gates as "America's richest 'nerd.' " By 1992, at age 36, he had passed Donald Trump, Ross Perot, and others to be listed as America's wealthiest person by *Forbes* magazine; at that time the value of his holdings had grown to an estimated \$6.3 billion. More recently his wealth surpassed \$40 billion. How did the free enterprise system help him to attain such phenomenal wealth?

After graduating from high school in Seattle in 1973, Gates went to Harvard. While there, he learned that the personal computer (PC) was in the development stage. He dropped out of school and threw himself completely into designing an *operating system* for the PC. The operating system is the program that coordinates the hardware and software of the computer. His system, MS-DOS (Microsoft Disk Operating System), was so good that IBM agreed to use it in their line of personal computers. With IBM setting the industry standard, other computer manufacturers quickly adopted MS-DOS as well. Later, MS-DOS was replaced by the Windows operating system, another of Microsoft's developments. Today more than 80 percent of all personal computers in the world use this system. Gates' firm, Microsoft, Inc., makes money on every computer sold with Windows as the operating system. In 1997, the firm recorded over \$10 billion in revenue and more than \$2 billion in net profit. It ranks third in size in the industry, behind IBM and Hewlett-Packard. Gates' personal holdings of some 300 million shares of common stock represent about a 25 percent ownership share of the company.

Microsoft also produces programs for word processing, spreadsheets, and a variety of other applications. One of Gates' ventures has been to purchase the electronic reproduction rights to thousands of art and photographic works from museums and libraries around the world. These will be used as part of his plan for interactive home entertainment systems.

With extremely hard work, a creative mind, and a willingness to take risks, Gates has demonstrated how the market rewards the successful entrepreneur. He was able to produce what consumers wanted at a price they were willing to pay; the result was that both he and they are better off! This is the essence of the free-market economic system. ■

## ECONOMICS AND DECISION MAKING

Where do the principles of microeconomics fit in the arena of managerial decision making? Nobel prize-winning economist Herbert Simon identifies the primary activities in decision making:[8]

1. Finding occasions for making decisions.
2. Identifying possible courses of action.
3. Evaluating the revenues and costs associated with each course of action.
4. Choosing that one course that best meets the goal or objective of the firm (i.e., that maximizes the value of the firm).

The primary role of managerial economics is evaluating the implications of alternative courses of action and choosing the best or optimal course of action from among those several alternatives.

[8] H. A. Simon, "The Decision-Making Process," in *The Executive as Decision-Maker: The New Science of Management Decisions* (New York: Harper & Row, 1960).

Decision making in this context implies the need for optimizing behavior. The marketing vice-president strives to maximize sales revenue, the production manager attempts to minimize cost or maximize production, and the division president's goal is to maximize profit. As discussed in an earlier section, these management targets are constrained by other parameters relating to that decision. For example, production costs might be minimized by producing nothing, but this would be inconsistent with the firm's goal of profit maximization. A more typical goal for the production manager would be to minimize cost, subject to producing a specified output rate, while the objective of the marketing vice-president would be to maximize sales, subject to a given advertising budget.

The essence of efficient and rational management is constrained optimization. Virtually all choices and decisions are subject to limitations, and this is where the tools of managerial economics are most useful. The manager who can achieve the most despite those constraints will be rewarded with a high salary, stock options, and the other perquisites usually associated with success.

Optimization principles of managerial economics are also important in the not-for-profit sectors of the economy. For example, universities strive to maximize the value of teaching and research outputs, subject to an annual budget constraint. Decisions about the level of tuition, the right mix of faculty and secretaries, and the balance between classrooms and laboratories will benefit from the correct application of the constrained optimization principles that are basic to economics.

**Key Concepts**

- Economic profit refers to revenues minus all relevant costs, both explicit and implicit.
- Profit plays two roles in a market economy: (1) Changes in profit signal producers to change the rate of production; and (2) profit is a reward to entrepreneurs for taking risks, being especially innovative in developing new products, and reducing production costs.
- Firms can earn economic profits because they have monopoly power in a market. In general, such profits are not socially useful.
- The primary decision-making role of managerial economics is in determining the optimal course of action where there are constraints imposed on the decision.

## MMARY

Managerial economics can be viewed as an application of that part of microeconomics that focuses on such topics as risk, demand, production, cost, pricing, and market structure. Understanding these principles will help to develop a rational decision-making perspective and will sharpen the analytical framework that the executive must bring to bear on managerial decisions.

Individuals and firms interact in both the product and the factor markets. Prices of outputs and inputs are determined in these markets and guide the decisions of all market participants. The firm is an entity that organizes factors of production in order to produce goods and services to meet the demands of consumers and other firms. In a market system, the interplay of individuals and firms is not subject to central control. The prices of both products and factors of production guide this interaction. Within firms, however, activity is directed by managers. Central control within the firm is advantageous because transactions and information costs are reduced. The size of the firm is limited because transaction costs within the firm will rise as the firm grows, and because management skill is limited.

It is assumed that the goal of the firm is to maximize the value of the firm or the present value of all future profits, defined as revenue less all costs, explicit and implicit. Opportunity costs such as the remuneration and interest that owners and managers have foregone on their labor and capital must be included as costs. Failure to account for these implicit costs may result in an inefficient allocation of resources. The objective of profit maximization is subject to legal, moral, contractual, financial, and technological constraints. Some economists argue that the firm's objective is a "satisfactory" level of profit rather than maximum profit. The principal-agent problem arises where the owner of a firm and the manager of that firm have different objectives. The problem can be solved by tying part of the manager's salary to profits and/or changes in the price of the firm's stock.

Profit plays two primary roles in the free-market system. First, it acts as a signal to producers to increase or decrease the rate of output, or to enter or leave an industry. Second, profit is a reward for entrepreneurial activity, including risk taking and innovation. In a competitive industry, economic profits tend to be transitory. The achievement of high profits by a firm usually results in other firms increasing their output of that product, thus reducing price and profit. Firms that have monopoly power may be able to earn above-normal profits over a longer period; such profit does not play a socially useful role in the economy.

A primary role of economics in management is in making optimizing decisions where constraints apply. The application of the principles of managerial economics will help managers ensure that resources are allocated efficiently within the firm, and that the firm makes appropriate reactions to changes in the economic environment.

## Discussion Questions

1-1. Explain how firms and individuals participate and interact in the product market and in the factor market.

1-2. Describe the difference between the accounting and the economic concept of profit. How might accounting practices be changed to make financial statements and reports more useful for managerial decision making?

1-3. Explain how the principles of free-market economics that guide interrelations among firms and individuals might guide pricing and resource allocation decisions within the large, multiplant firm.

1-4. Why is it important to state a managerial objective? Could the assumption that the managers' objective is profit maximization be useful even if their real objective is maximizing market share or their salaries?

1-5. What might be the objective or objectives of each of the following nonprofit institutions?

a. The college of business at a major state university.

b. A municipal police department.

c. The emergency room of a hospital.
d. A museum.

1-6. Explain the role of profit in a free-market economic system. How can such a small share of national income (about 15 percent) be such an important determinant of resource allocation?

1-7. Provide examples of managerial decisions that might reduce profits for the next few years, but would increase the value of the firm. Explain.

1-8. Some argue that businesses need to be "socially responsible" (e.g., reduce pollution, employ more minority workers, cease buying from and selling to countries that are out of political favor, sell at lower prices to low-income people, etc.). Evaluate the effects on the firm of being socially responsible in a competitive environment. Can one firm in a market afford to be socially responsible if its competitors are not? Explain. Would you invest in a company that devoted a significant part of its resources to being socially responsible? Why?

1-9. Milton Friedman has argued that the realism of a model is not as important as its predictive ability. That is, a model of economic behavior that appears to be unrealistic but predicts well is superior to a model that seems more realistic but does not predict well. Do you agree? Why or why not?

1-10. In some large businesses, division managers must report profits monthly and, in some cases, more often. Some managers claim that the pressure to report favorable results continually on a month-to-month basis causes them to make decisions that will enhance short-term profit at the expense of profit in the long run. What actions might a manager take that would increase profit in the short run but that would reduce shareholder value?

1-11. A firm's profit has been at approximately $10 million per year in each of the last five years. The company has two million shares of stock outstanding, and the market price has been about $60 per share. The top management officials have been receiving only a salary during this period. Describe how a combined bonus and stock-option plan could be developed to provide these executives with greater incentive to increase profit.

## Problems

1-1. A recent engineering graduate turns down a job offer at $30,000 per year to start his own business. He will invest $50,000 of his own money, which has been in a bank account earning 7 percent interest per year. He also plans to use a building he owns that has been rented for $1,500 per month. Revenue in the new business during the first year was $107,000, while other expenses were

| | |
|---|---|
| Advertising | $ 5,000 |
| Rent | 10,000 |
| Taxes | 5,000 |
| Employees' salaries | 40,000 |
| Supplies | 5,000 |

Prepare two income statements, one using the traditional accounting approach and one using the opportunity cost approach to determine profit.

1-2. Tempo Electronics, Inc., has an inventory of 5,000 unique electronic chips originally purchased at $2.50 each; their market value is now $5 each. The production department has proposed to use these by putting each one together with $6 worth of labor and other materials to produce a wristwatch that would be sold for $10. Should that proposal be implemented? Explain.

1-3. Smith, a college sophomore, generally spends his summers working on the university maintenance crew at a wage rate of $6.00 per hour for a 40-hour week. Overtime work is always available at an hourly rate of 1.5 times the regular wage rate. For the coming summer, he has been offered the pizza stand concession at the Student Union building, which would have to be open 10 hours per day, six days a week. He estimates that he can sell 100 pizzas a week at $6.00 each. The production cost of each pizza is $2.00 and the rent on the stand is $150 per week. Should Smith take the pizza concession? Explain.

1-4. An executive's employment contract calls for a salary of $400,000 per year, a bonus equal to 2 percent of profits in excess of $10,000,000, and an option to buy 5,000 shares of common stock at a price of $50 per share. The market price of the stock is $70 per share, and the firm's profits for the current year are $12,000,000. Assuming the executive exercises the stock option and then sells the stock, what is the total compensation for the year?

1-5. A manufacturer of personal computers has an inventory of 10,000 back-up storage drives that sold for $100 per unit last year. The current market price of these drives is now $70 per unit, By adding one of these drives to their stock of personal computers, the price of each computer is increased by $80 per unit. Should the driver be added? What is the opportunity cost of these drivers? Explain.

# Appendix

# *Present-Value Analysis*

Many transactions involve making or receiving cash payments at various future dates. A person who takes out a mortgage loan trades a promise to make monthly payments for thirty years for a large amount of cash now to pay for a home. A person injured in an automobile accident accepts an insurance company's settlement of $1,000 per month for life as compensation for the damage associated with that injury. High-priced professional athletes offer their skills for multiyear, no-cut contracts. In all these cases, concepts relating to the time value of money are required to make sound decisions. The time value of money refers to the fact that a dollar to be received in the future is not worth a dollar today. Therefore, it is necessary to have techniques for measuring the value today (i.e., the present value) of dollars to be received or paid at different points in the future. This section outlines the approach to analyzing problems that involve payment and/or receipt of money at one or more points in time.

Understanding the following terms is essential to applying time value of money principles:

*Annuity:* a series of equal payments per period for a specified length of time. For example, the repayment of a loan by making forty-eight monthly payments of $200 each is a form of annuity.

*Amount:* a specific number of dollars to be paid or received on a specified date.

*Present Value:* the value today of an amount or an annuity, taking into consideration that interest can be earned.

## PRESENT VALUE OF AN AMOUNT

The basic equation for the present value ($PV$) of an amount $S$ is

$$PV = S\left[\frac{1}{(1+i)^n}\right] \qquad (1A\text{–}1)$$

The bracketed term

$$\left[\frac{1}{(1+i)^n}\right]$$

is the present value of $1 in $n$ periods if the interest rate is $i$ percent. It is called the *present-value interest factor* ($PVIF_{i,n}$).

As an example, what is the present value of $1,080 in 1 year if the interest rate is 8 percent per year? Substituting $S = 1{,}080$, $i = 0.08$ (i.e., the decimal equivalent of 8 percent), and $n = 1$ in equation (1A-1) yields

$$PV = \$1{,}080\left[\frac{1}{(1.08)^1}\right] = \$1{,}000$$

Note that \$1,000 would increase to \$1,080 in one year at 8 percent interest. Thus, the present-value concept explicitly takes account of the potential interest that could be earned.

Consider another problem. What is the present value of \$100,000 to be received at the end of 10 years if the interest rate is 10 percent? The problem can be expressed as

$$PV = 100{,}000\left[\frac{1}{(1.10)^{10}}\right]$$

The present-value interest factors ($PVIF_{i,n}$) for a range of interest rates and periods are given in Table I at the end of the book. Part of that table is reproduced as Table 1A.1.

**TABLE 1A.1 Selected Present-Value Interest Factors**

$$PVIF_{i,n} = \frac{1}{(1+i)^n}$$

| | *Interest Rate* | | |
|---|---|---|---|
| ***Periods*** | ***8%*** | ***10%*** | ***12%*** |
| 1 | 0.9259 | 0.9091 | 0.8929 |
| 2 | 0.8573 | 0.8264 | 0.7972 |
| 3 | 0.7938 | 0.7513 | 0.7118 |
| 4 | 0.7350 | 0.6830 | 0.6355 |
| 6 | 0.6302 | 0.5645 | 0.5066 |
| 8 | 0.5403 | 0.4665 | 0.4039 |
| 10 | 0.4632 | 0.3855 | 0.3220 |

By reading down the "10%" column in Table 1A.1 to the row for $n = 10$ periods, the factor 0.3855 is found. This is the present value of \$1 in 10 years at 10 percent interest. Multiplying this factor by \$100,000, we obtain

$$PV = 100{,}000(0.3855) = \$38{,}550$$

If 10 percent is the appropriate interest rate, \$100,000 in 10 years is equivalent to \$38,550 today.

The process of reducing a future amount to its present value is often referred to as *discounting* because the present value is always less than the future amount. In this context, the interest rate used in present-value problems is generally referred to as a *discount rate.*

Note that the present-value factors decrease as the number of periods increases and as the interest rate increases. Because the interest rate is in the denominator of the present-value equation, there is an inverse relationship between the present value and the interest rate. Furthermore, the longer the period of time before an amount is paid, the lower the present value of any amount.

## PRESENT VALUE OF AN ANNUITY

An *annuity* has been defined as a series of periodic equal payments. Although the term is often thought of in terms of a retirement pension, there are many other examples of annuities. The repayment schedule for a mortgage loan is an annuity. A father's agreement to send his son \$200 each month while he is in college is another example. Usually, the number of periods is specified, but not always. Sometimes retirement benefits are paid monthly as long as a person is alive. In other cases, the annuity is paid forever and is called a *perpetuity*.

It must be emphasized that the strict definition of an annuity implies equal payments. A contract to make 20 annual payments, which increase each year by, say, 10 percent, would not be an annuity. As some financial arrangements provide for payments with periodic increases, care must be taken not to apply an annuity formula if the flow of payments is not a true annuity.

The present value of an annuity can be thought of as the sum of the present values of each of several amounts. Consider an annuity of three \$100 payments at the end of each of the next three years at 10 percent interest. The present value of each payment is

$$PV_1 = 100\frac{1}{1.10}$$

$$PV_2 = 100\frac{1}{1.10^2}$$

$$PV_3 = 100\frac{1}{1.10^3}$$

and the sum of these would be

$$PV = 100\frac{1}{1.10} + 100\frac{1}{(1.10)^2} + 100\frac{1}{(1.10)^3}$$

or

$$PV = 100\left[\frac{1}{1.10} + \frac{1}{(1.10)^2} + \frac{1}{(1.10)^3}\right]$$

Substituting the appropriate present-value interest factors from Table 1A.1 and multiplying yields, the present value of this annuity is

$$PV = 100(0.9091 + 0.8264 + 0.7513) = 100(2.4868) = 248.68$$

Although this approach works, it clearly would be cumbersome for annuities of more than a few periods. For example, consider using this method to find the present value of a monthly payment for forty years if the monthly interest rate is 1 percent. That would require evaluating the present value of each of 480 amounts!

In general, the formula for the present value of an annuity of $A$ dollars per period for $n$ periods and a discount rate of $i$ is

$$PV = A\frac{1}{(1+i)} + A\frac{1}{(1+i)^2} + \cdots + A\frac{1}{(1+i)^n}$$

This can be written as

$$PV = A\sum_{t=1}^{n}\left[\frac{1}{(1+i)^t}\right] \tag{1A-2}$$

Equation (1A-2) is the general equation for the present value of an annuity. Recall that $\sum_{t=1}^{n}$ means the sum of $n$ separate components, the first where $t = 1$, the second where $t = 2$, and so on, to $t = n$.

The term

$$\sum_{t=1}^{n}\left[\frac{1}{(1+i)^t}\right]$$

is called the *present-value annuity factor* ($PVAF_{i,n}$). It is the present value of an annuity of \$1 per period for $n$ periods at a discount rate of $i$ percent. Table II at the end of the book provides these factors for a variety of interest rates and periods. Part of that table is reproduced as Table 1A.2 here.

**TABLE 1A.2 Selected Present-Value [illegible] Factors**

$$PVAF_{i,n} = \sum_{t=1}^{n}\frac{1}{(1+i)^t}$$

| | *Interest Rate* | | |
|---|---|---|---|
| ***Periods*** | ***1%*** | ***2%*** | ***3%*** |
| 12 | 11.2551 | 10.5753 | 9.9540 |
| 24 | 21.2434 | 18.9139 | 16.9355 |
| 30 | 25.8077 | 22.3965 | 19.6004 |

Consider the present value of an annuity of \$3,522 per month for thirty months with an interest rate of 1 percent per month. Note that this problem differs from the first two examples because it considers monthly (not annual) payments and a monthly discount rate. This should not be confusing. The general problem refers to $n$ periods and a discount rate of $i$ percent *per period.* As long as the length of period (i.e., month, year, etc.) and the interest rate for that period correspond, the approach is straightforward. For example, if the periods are years, the interest rate must be a yearly rate.

The equation for the present value of this monthly annuity is

$$PV = 3{,}522\left[\sum_{t=1}^{30}\left(\frac{1}{1.01}\right)^t\right]$$

The factor

$$\sum_{t=1}^{30}\frac{1}{(1.01)^t}$$

is the present value of an annuity of \$1. From Table 1A.2, that value is 25.8077. Substituting that value and multiplying gives $PV = \$3{,}522(25.8077) = \$90{,}895$. This means that an amount of $90,895 invested at an interest rate of 1 percent per month would be just adequate to make thirty monthly payments of $3,522.

A similar problem might be stated in the following way. What is the present value of a series of 30 monthly payments of $150 if the interest rate is 24 percent per year? Note that the payments are monthly, but the interest rate is stated as an annual rate. By dividing the annual interest rate by 12, the appropriate monthly rate is found (i.e., 24/12 = 2). Thus the solution would be

$$PV = 150\left[\sum_{t=1}^{30}\left(\frac{1}{1.02}\right)^t\right] = 150(22.3965) = \$3,359$$

With the advent of financial calculators, there is less need for actually using the present-value equations and the tabular data on present-value interest and annuity factors. However, it is essential that the principles underlying these financial calculations be understood.

## Example Solomon Keith and the New York Lottery

Solomon Keith, a 55-year-old bank custodian, won the $5 million New York lottery in 1987, but unfortunately died in an auto accident 15 months later after collecting only the first two installments of $240,245 each. As the *Wall Street Journal* put it: "...his number came up not once but twice." His estate collected three more payments, leaving 16 annual payments to be made.*

Primarily because of income taxes, estate taxes, and the administrative costs associated with the estate, the remaining payments were put up for bid. The Prudential Life Insurance Company made the winning bid of $2,075,000 (of which $700,000 was used to pay estate taxes, $550,000 went to attorneys and the administrator of the estate, and the remaining $825,000 was divided among Keith's nine relatives).

What discount rate did Prudential use to come up with their winning bid?

**Solution** The discount rate is found by using the present value of an annuity formula

$$PV = \sum_{t=1}^{n}\frac{A}{(1+i)^t}$$

and making appropriate substitutions, that is,

$$2,075,000 = 240,245\left[\sum_{t=1}^{16}\frac{1}{(1+i)^t}\right]$$

solving for the present-value annuity factor

$$(PVAF_{i,n}) = \frac{2,075,000}{240,245} = 8.6370$$

and then using Table II (or a financial calculator) to determine that $i = 0.0839$ (i.e., a discount rate of 8.39 percent was used.)

*See *Wall Street Journal*, June 30, 1992, and the *Salt Lake Tribune*, June 3, 1992.

**Key Concept**

Given the interest rate per period ($i$), the number of periods ($n$), and the amount ($S$) or the annuity payment ($A$), there are two basic present-value problems:

1. Present Value of an Amount: $PV = S\left[\frac{1}{(1+i)^n}\right]$

$PV$ is the present value of an amount $S$ to be received (or paid) in $n$ periods if the interest rate is $i$ percent per period. The term $1/(1+i)^n$ is the present-value interest factor ($PVIF_{i,n}$).

2. Present Value of an Annuity: $PV = A\left[\sum_{t=1}^{n}\frac{1}{(1+i)^t}\right]$

Here $PV$ is the present value of an annuity of $A$ per period paid at the end of each of $n$ periods if the interest rate is $i$ percent per period. The term $\sum_{t=1}^{n}[1/(1+i)^t]$ is the present value annuity factor $PVAF_{i,n}$.

## Problems

1A-1. Robert Ryan, general manager of the Chicago Stars professional football team, is currently negotiating a new contract with Ronnie Smith, the team's star running back. Under league rules, Smith is now a free agent, which means that he is free to negotiate a contract with any other team in the league. Smith has presented Ryan with a final contract demand consisting of alternatives for a five-year contract. If Ryan does not agree to one of these, Smith will sign with another team. The alternative contract demands are

a. A $2,000,000 bonus payment immediately, a payment of $500,000 at the end of each of the next five years, and a deferred payment of $1,000,000 at the end of the fifth year of the contract.

b. A $500,000 bonus payment now, payments of $300,000 at the end of each of the next five years, and deferred payments of $200,000 each payable at the end of years 11 through 20.

Ryan has determined that Smith's value to the team over the next five years is about $3,000,000 (in terms of the present value of additional revenue from gate receipts and television discounted at 12 percent per year). Should Ryan accept one of Smith's contract demands, and if so, which one? Explain fully.

1A-2. A rich uncle gives you the choice of one of the following legacies:

a. $15,000 each year for the next 12 years.

b. $13,000 each year for the next 18 years.

c. $11,000 each year for the next 12 years plus a lump-sum payment of $81,000 at the end of the 18th year.

Which would you take and why? Assume that the appropriate discount rate is 10 percent and all amounts would be received at the end of the year.

1A-3. Mity-Lite, Inc., a manufacturer of plastic tables for institutional use, is considering a capital spending program involving annual expenditures of $100,000 for each of the next five years. The firm estimates that its annual profit of $100,000 would increase by 50 percent when the capital program was completed. Assuming the firm has a 20-year life and the appropriate interest rate is 12 percent, should the capital spending program be implemented?

1A-4. The Smith Construction Company borrows the entire cost of a new dump truck. The loan has an annual interest rate of 12 percent and calls for monthly payments of $1,000 over a five-year period. What is the cost of the truck?

1A-5. A couple borrows $10,000 to buy a car. The loan agreement specifies that monthly payments are to be made for four years. The annual interest rate is 12 percent. Determine the monthly payment.

1A-6. A firm develops a new product that will add $50,000 to profit each year for five years. If the discount rate is 10 percent per year, how much will this new product add to shareholder value?

1A-7. It is estimated that the annual after-tax profits of the Microwave Corporation will be $500,000 per year for each of the next 30 years. Given a discount rate of 14 percent, what is the value of this firm?

1A-8. A loan agreement specifies that payments of $133.33 are to be made each month for five years. The annual interest rate specified is 6 percent. What is the amount of the loan?

1A-9. An executive's three-year employment contract calls for an annual salary of $600,000 (paid in monthly installments) for the first year with a 10 percent increase in the salary each year. What is the present value of this contract? The appropriate discount rate is 8 percent per year.

1A-10. The HAL Computer Corporation is considering an increase in its annual advertising expenditures from $10 million to $15 million for a five-year period (i.e., in years 1 to 5). The marketing department estimates that the increased advertising will increase profits by $4 million in years 3 to 7 and by $3 million in years 8 to 10, after which profits will return to the level they were at prior to the new program. If the firm uses a discount rate of 12 percent, will the proposed advertising program increase shareholder value?

1A-11. Lightco Inc., has just replaced its production machinery with new, more efficient equipment that will reduce annual production costs by $5 million per year. The cost of the new machines is $28.8 million. If the firm uses a discount rate of 8 percent per year, how many years will it take for the firm to recoup its investment?

1A-12. A state department of transportation is considering replacement of a bridge at a cost of $20 million. The life of the new bridge is 30 years, and it is estimated that improved safety and reduced congestion would be valued by bridge users at $3 million per year. The department uses an interest rate of 10 percent to evaluate capital projects.

a. Should this proposal be implemented?

b. If not, what is the maximum cost of the new bridge that could be justified by the benefits to users?

1A-13. On Monday, Smith borrows $100,000 at 8 percent annual interest with repayment to be made in monthly installments for 30 years. Overnight, the market interest rate increases to 12 percent. The lender, anxious to get its money back from Smith so that it can be loaned out at the higher interest rate, makes Smith the following offer: If she will double her monthly payment (thus significantly reducing the term of the loan), the lender will reduce the interest rate on Smith's loan to 6 percent per year. Should Smith take this offer? Explain. (Hint: Smith's opportunity cost is 12 percent per year. That is, she can invest at the same rate that the bank can lend.)

1A-14. A corporate bond specifies that interest of $400 is to be paid every six months for 20 years, and then the $1,000 principal or face value of the bond also will be paid. The market interest rate for bonds of this quality is 10 percent per year. What is the market value of this bond?

1A-15. A pedestrian is seriously injured when hit by an automobile and files a lawsuit against the driver. The driver's insurance company offers the victim a structured settlement, which the company announces to be worth $2,000,000. The offer consists of the following components:

*i.* $1,000 per month for 360 months;

*ii.* one-time payments of $100,000, $200,000, and $300,000 at the end of 35, 40, and 45 years, respectively; and

*iii.* a final payment of $1,000,000 at the end of the 50th year.

a. Given that the market interest rate is 12 percent per year, evaluate this offer.

b. Develop a structured settlement that has a true present value of $2,000,000.

# CHAPTER 2

# Basic Training

## PREVIEW

Just as the skilled craftsman needs tools to build a new home, the manager needs tools to assist in making decisions that will ultimately maximize the value of the firm. Increasingly, these tools are of the quantitative sort. Algebra, statistics, linear programming, and, to some extent, calculus are used in all the functional areas of business today. This is especially true for decision makers in production and finance, but even those in marketing and business law increasingly find these tools of value.

Many management decisions fall into the category of optimization problems. Optimization refers to finding the best way to allocate resources given an objective function. For example, a production manager may seek to maximize production for given inputs of capital and labor or to minimize the cost of producing a specified rate of output. In contrast, the president of the firm, who has a broader perspective, may want to organize the firm's labor and capital resources in order to maximize profit and shareholder value. Terms such as *minimize* and *maximize* imply an optimization problem. Sometimes managers seek optimal decisions by trial and error or other informal methods. In many situations, however, data on the firm and its market can be combined with quantitative analytical tools to scientifically determine the optimal management strategy.

The purpose of this chapter is to introduce the basic quantitative tools that are commonly used in managerial decision making and to demonstrate their applications. In the first section, functional relationships are discussed, and a special set of functions—total, average, and marginal—is developed and applied to problems in managerial economics. Next, the nature and structure of an economic model is discussed. Typically, such models include several equations and conditions that constrain the nature of these functions. In the last section, basic principles of probability are developed, with emphasis on measures of central tendency and dispersion for probability distributions.

In addition, a review of the elements of calculus and their application to economics is included as an appendix. While all of the important concepts in managerial economics can be understood without using calculus, its use helps to understand important relationships and allows certain problems to be solved.

## FUNCTIONAL RELATIONSHIPS: TOTAL, AVERAGE, AND MARGINAL

In mathematics, an equation of the form

$$y = f(x) \tag{2-1}$$

is read "$y$ is a function of $x$." This means that the value of $y$ depends on the value of $x$ in a systematic way and that there is a unique value of $y$ for each value of $x$. Usually the variable on the left-hand side of the equation, $y$ in this case, is called the *dependent variable*. The variable on the right-hand side is called the *independent variable*.

Of course, $y$ may depend on two or more independent variables, such as

$$y = g(x, z) \tag{2-2}$$

or

$$y = h(x_1, x_2, \ldots, x_n) \tag{2-3}$$

The letters $f$, $g$, and $h$ have no specific meaning other than to indicate that a functional relationship exists between $y$ and the independent variable or variables. Equation (2-1) is referred to as a univariate function because it has only one independent variable. Equations (2-2) and (2-3) are multivariate functions.

In many economic models, a special set of functional relationships called total, average, and marginal functions is used. Such functions are involved in the theory of demand, cost, production, and market structure. A basic command of these concepts is essential to understanding the principles of managerial economics.

The following production example will help in understanding these relationships. Suppose that there is a small building containing four machines and a stock of raw materials ready to be processed. Ten equally skilled and diligent workers are lined up outside ready to go to work in this factory. If there are no workers, output will be zero. As workers are added, output increases. The total amount of output associated with a particular input of labor working with those four machines is called the *total product of labor.* For example, the total product of one worker might be two units of output. As the labor input changes, so does total output. An example of a total product schedule is shown in the first two columns of Table 2.1.

**TABLE 2.1 Total, Average, and Marginal Product of Labor Schedules**

| *Number of Workers (L)* | *Total Product (Q)* | *Average Product (AP)* | *Marginal Product (MP)* |
|---|---|---|---|
| 0 | 0 | — | |
| 1 | 2 | 2.0 | 2 |
| 2 | 5 | 2.5 | 3 |
| 3 | 9 | 3.0 | 4 |
| 4 | 14 | 3.5 | 5 |
| 5 | 22 | 4.4 | 8 |
| 6 | 40 | 6.7 | 18 |
| 7 | 57 | 8.1 | 17 |
| 8 | 63 | 7.9 | 6 |
| 9 | 64 | 7.1 | 1 |
| 10 | 63 | 6.3 | −1 |

As just indicated, one person working alone in this factory produces two units of output. Adding a second person and organizing the production system so that the workers complement each other results in total product increasing to five units. The first worker is associated with a two-unit increase in output; having two workers instead of one will increase output by three units; three workers will increase output by four units; and so on. The change in output associated with a one-unit change in workers is called the *marginal product of labor.* Using the Greek capital letter delta ($\Delta$) to indicate a change, the *marginal product function* ($MP$) can be defined as

$$MP = \frac{\Delta Q}{\Delta L} \tag{2-4}$$

where $Q$ represents output and $L$ represents the input of labor.[1]

Note in Table 2.1 that for the first six workers, the marginal product increases as the rate of labor input increases. However, marginal product declines thereafter. That is, the sixth worker adds 18 units to output, but the seventh adds only 17 units and the eighth worker only 6. Finally, having 10 workers actually causes output to decline; the marginal product associated with the tenth worker is negative. The point has been reached where there are too many workers in this plant. Perhaps they are getting in each other's way; in any event, the presence of 10 workers has thwarted the achievement of efficient production.[2]

The *average product of labor function* ($AP$) measures the average output per unit of labor used. Average product is found by dividing total product by labor input. That is,

$$AP = \frac{TP}{L} \tag{2-5}$$

The total product function is plotted in Figure 2.1a, and the average and marginal product functions are shown in Figure 2.1b.[3] There are important relationships among the three functions that are true for all total, average, and marginal functions. First, the value of the average function at any point along that curve is equal to the slope of a ray drawn from the origin to the total function at the corresponding point. For example, from Table 2.1 it is known that the average product of six workers is 6.7. Thus, the slope of a line ($OA$) drawn from the origin to point $A$ on the total product function has a slope of 6.7. Similarly, the average product of seven workers is 8.1; this is equal to the slope of a line ($OB$) drawn from the origin to point $B$ on the total product function.

---

[1]The marginal product function is the slope of the total product function. Slope can be defined as the change in the dependent variable divided by the change in the independent variable. Consider the function $y = 10 + 3x$. A one-unit change in $x$, say from $x = 5$ to $x = 6$, is associated with a three-unit change in $y$, from $y = 25$ to $y = 28$. Thus the change in $y$ (i.e., $\Delta y$) is 3, and the change in $x$ (i.e., $\Delta x$) is 1, so the slope is $\Delta y/\Delta x = 3/1 = 3$.

[2]That marginal product must ultimately decline as workers are added is clear in this example. The building has a finite amount of floor space. The change in output should be positive if one or two workers are added. But for some number of workers, output will be zero as workers fight for a place to stand rather than produce output. Therefore, beyond some rate of labor input, marginal product must decline.

[3]All three of the functions are drawn with graphs having the same units on both axes (i.e., output on the vertical axis and number of workers on the horizontal axis). Therefore, all three could have been included in the same graph. However, the range of total product (0 to 64) is so much greater than the range of average and marginal product that it is more descriptive to use two graphs. Note that the height of each interval on the vertical axis in Figure 2.1b is considerably greater than in Figure 2.1a. This allows the relationships among all three functions to be seen more clearly than if they were all drawn in the same graph. Note also that the horizontal axis in both graphs is exactly the same. This allows comparison of points on one graph that correspond to points on the other.

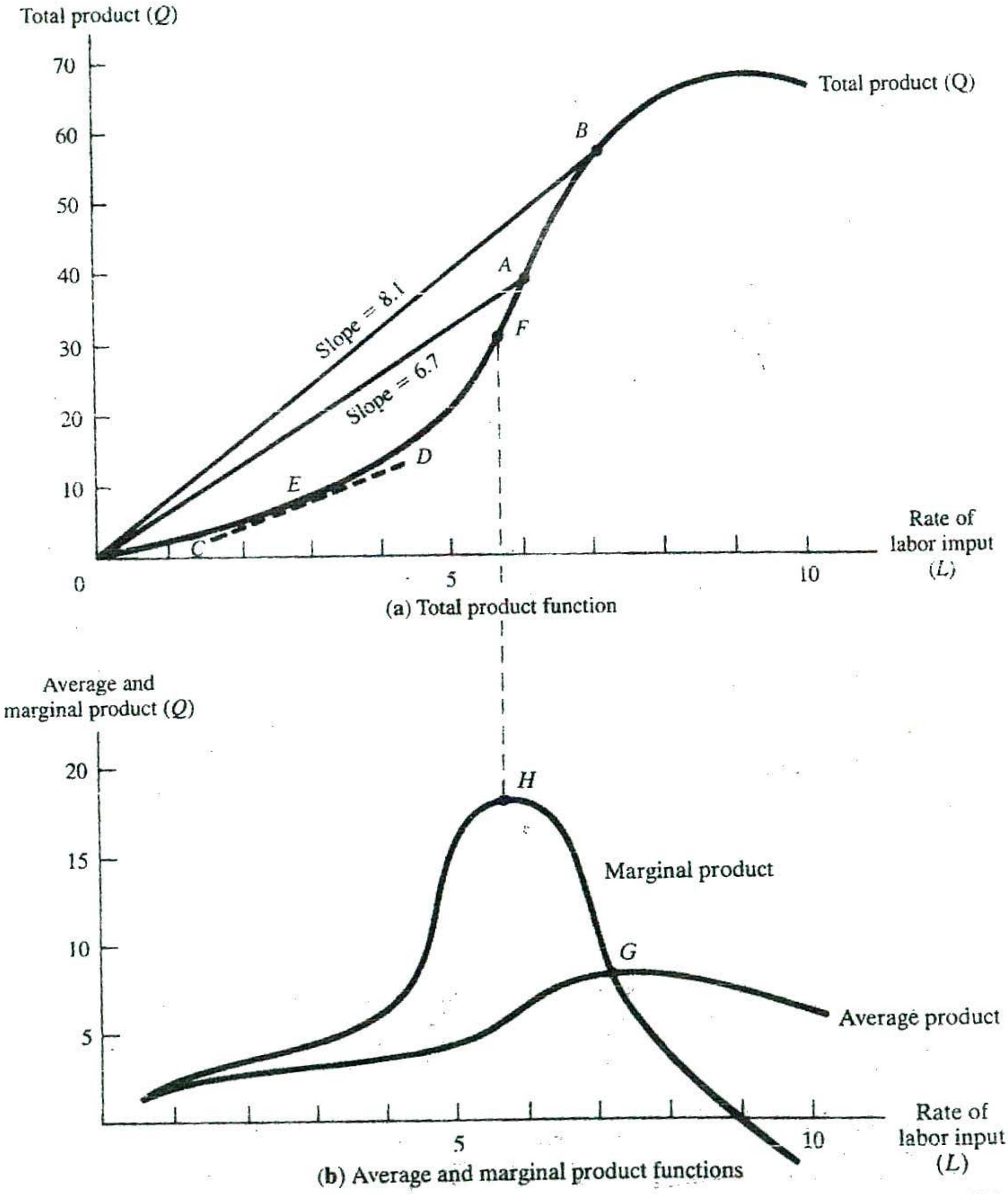



**FIGURE 2.1 The Product Functions**

Another key relationship is that the value of the marginal function is equal to the slope of the line drawn tangent to the total function at a corresponding point. For example, the slope of the dashed line *CD*, which is drawn tangent to the total function at *E*, is about 4. This means that the marginal product corresponding to this point (i.e., between 2 and 3 units of labor) is 4.

Point *F* on the total function is called an *inflection point*. To the left of point *F*, the total function is increasing at an increasing rate; to the right of *F*, the total function is increasing but at a decreasing rate. Note that this inflection point, corresponding to about six workers, occurs at the point where the marginal product function is at a maximum (i.e., point *H* in Figure 2.1b).

If the marginal and average functions intersect, that point of intersection will be at the minimum or maximum point on the average function. In Figure 2.1b, the intersection occurs at point *G*, which is the maximum point of the average product function. The logic of this relationship is quite straightforward. Suppose that the average product for two workers is 2.5 and the marginal product of three workers is more than that average, say, 4. Thus the average product for three workers must increase to 3. If the additional output associated with hiring another worker is above the previous average product, the average product must be increasing.

Conversely, suppose that eight workers average 7.9 units of output per period and that the marginal product of nine workers is only 1. This will cause average product to fall to 7.1. If an additional worker adds less to total output than the average product prior to that addition, average product must fall.

This logic leads to the following conclusion. For any set of average and marginal functions, if the marginal function is greater than the average function, the average must be rising. If marginal is below average, the average must be falling. This implies that the intersection of the two functions must occur where the average function is at a maximum or minimum. As will be shown in later chapters, this result is important in some managerial economics problems.

If the marginal function is positive, the total function must be rising. Note that it does not matter whether the marginal function is increasing or decreasing, as long as it is positive. Conversely, if the marginal function is negative, the total function must be declining. Again, it does not matter if marginal is rising or falling. If marginal is negative, the total function will be declining. For the data in Table 2.1 and Figure 2.1, the marginal product is positive for the first nine workers. It declines after the sixth worker but remains positive through the ninth. Note that total product increases until the tenth worker is added. The marginal product of the tenth worker is $-1$, and this negative marginal product is associated with a decline in total product.

Because the total function increases as long as the marginal function is positive and decreases when marginal is negative, it follows that total product is at a maximum when the marginal function is zero. In Figure 2.1 the maximum of the total product function occurs at nine workers. This point corresponds to the point where the marginal function intersects the horizontal axis, that is, where marginal changes from being positive to negative.

An understanding of total, average, and marginal relations is an important foundation for the effective study of managerial economics. Terms such as *average cost*, *marginal product*, and *total revenue* are integral parts of the manager's vocabulary, and the associated principles are some of the manager's most powerful tools.

# *Case Study*

## Decision Making in the Public Sector: Marginal Analysis and Automobile Safety

Suppose a specific automobile model has so many engineering defects that the risk of having an accident is greatly increased. Although the car is no longer being manufactured, vehicles already on the road are expected to cause 10 deaths during the next year. For simplicity, assume that the engineering defects are so severe that none of the cars will be running after one year. The number of deaths can be reduced if the manufacturer recalls the cars and corrects some or all of the defects. Clearly, the more defects that are corrected, the greater will be the reduction in the number of deaths.

As shown in the table, the marginal cost of repairing and modifying these cars in order to reduce the number of deaths from 10 to nine is $200,000, but marginal cost rises as the number of deaths decreases. For example, the marginal cost of going from one death to no deaths is $2,000,000. This cost rises because some of the defects are easily corrected, whereas others require a substantial modification of the vehicle.

Suppose that the National Highway Traffic Safety Administration, a federal agency responsible for regulating automobile safety, uses marginal principles to make recall decisions and that the agency has determined that each life saved is worth $800,000 to society. That is, $800,000 is the marginal benefit associated with saving a life.

The benefits and costs of the recall/repair program shown in the table can be compared to make a decision. The marginal cost of saving the first life is $200,000 and the marginal benefit is $800,000. Thus it is clear that the manufacturer should be required to take some action. However, it is equally clear that it would be inefficient to attempt to eliminate all defects. For example, the marginal cost of going from one death to zero deaths is $2,000,000 but the marginal benefit is only $800,000.

| *Deaths* | *Marginal Cost of Death Prevention* | *Marginal Benefit of a Life Saved* | *Total Net Benefits* |
|---|---|---|---|
| 10 | $ 0 | | |
| 9 | 200,000 | $800,000 | $ 600,000 |
| 8 | 400,000 | 800,000 | 1,000,000 |
| 7 | 600,000 | 800,000 | 1,200,000 |
| 6 | 800,000 | 800,000 | 1,200,000 |
| 5 | 1,000,000 | 800,000 | 1,000,000 |
| 4 | 1,200,000 | 800,000 | 600,000 |
| 3 | 1,400,000 | 800,000 | 0 |
| 2 | 1,600,000 | 800,000 | −800,000 |
| 1 | 1,800,000 | 800,000 | −1,800,000 |
| 0 | 2,000,000 | 800.000 | −3,000,000 |

The recall program should be designed to reduce those defects (and associated deaths) to the point where marginal benefits equal marginal costs. In this example, that equality occurs at six deaths. Note that total net benefits (a *net benefit* is defined as marginal benefit minus marginal cost) are maximized ($1,200,000) at six deaths. Total benefits are $3,200,000 (i.e., four lives saved multiplied by $800,000 per life) and total costs are $2,000,000. To require additional repairs to reduce the number of deaths below six is not socially efficient because the costs exceed the benefits. ■

---

**Key Concepts**

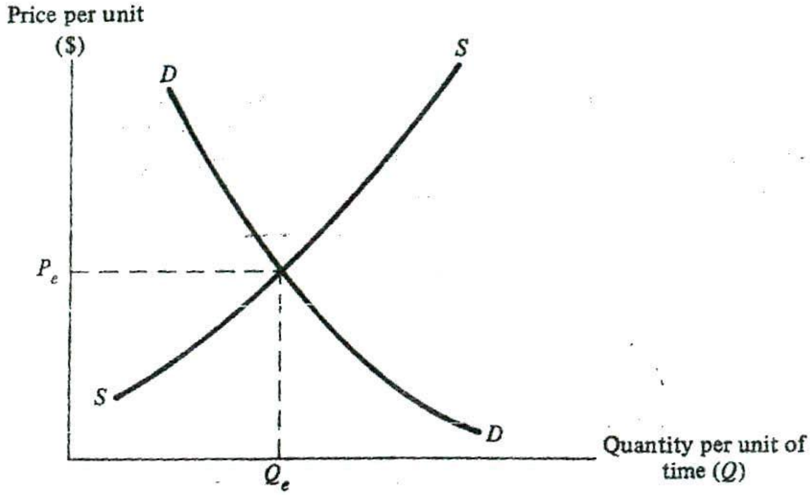
- A functional relationship of the form

$$y = f(x_1, x_2, \ldots, x_n)$$

means that there is a systematic relationship between the dependent variable $y$ and the independent variables $x_1, x_2, \ldots, x_n$ and that there is a unique value of $y$ for any set of values of the independent variables.
- For any total function (e.g., total product, total revenue, etc.) there is an associated marginal function and average function.
- The key relationships among the total, average, and marginal functions are
  1. The value of the average function at any point is the slope of a ray drawn from the origin to the total function at that point.
  2. The value of the marginal function at any point is the slope of a line drawn tangent to the total function at that point.
  3. The marginal function will intersect the average function at either a minimum or a maximum point of the average function.
  4. If the marginal function is positive, the total function will be increasing. If the marginal function is negative, the total function will be decreasing.
  5. The total function reaches a maximum or minimum when the marginal function equals zero.

## ECONOMIC MODELS

In the aerospace industry, small model airplanes are flown in wind tunnels to test the flight characteristics of full-sized planes having the same characteristics. In economics, graphs and/or equations are used to explain economic relationships and phenomena and to predict the effects of changes in such economic parameters as prices, wage rates, and the price of capital. Although such models are abstractions from reality and may seem unrealistic, they are useful in studying the way an economic system works. Just as it is not sound practice to build a new jet aircraft before a model is used to test its flight characteristics, an economic decision should not be made without having first analyzed its implications by using an economic model.




An economic model usually consists of several related functions, some restrictions on one or more of the coefficients of these functions, and equilibrium conditions. Recall the concepts of supply and demand from introductory economics. As shown in Figure 2.2, the demand curve ($DD$) slopes downward from left to right and shows the quantity of output that consumers are willing and able to buy at each price. The negative slope implies that a larger quantity is demanded at lower prices than at higher prices. The supply curve ($SS$) shows the amount that firms will produce and offer for sale at each price. This curve has a positive slope because firms will supply a larger quantity at higher prices than at lower prices.[4]

Equilibrium in a market exists when the quantity demanded equals the quantity supplied. This is shown graphically as the intersection of the demand and supply functions in Figure 2.2. In this example, $P_e$ and $Q_e$ are the equilibrium price and quantity, respectively. In equilibrium there is no incentive for buyers or sellers to change price or the quantity. At point $\{P_e, Q_e\}$, buyers' demands are met exactly and suppliers are selling exactly the number of units they desire to sell at that price.

The preceding example is an economic model depicted graphically. A simple algebraic model can be used to describe exactly the same economic phenomenon. The quantity demanded ($Q_d$) and the quantity supplied ($Q_s$) are both functions of price. That is,

$$Q_d = f(P) \text{ and } Q_s = g(P)$$

Suppose that the demand function is

$$Q_d = a + bP \text{ where } b < 0 \tag{2-6}$$

and the supply function is

$$Q_s = c + dP \text{ where } d > 0 \tag{2-7}$$

[4]The principles of demand and supply are discussed in greater detail in chapters 3 and 9.

The restrictions on the parameters (i.e., $b < 0$ and $d > 0$) simply mean that the demand curve must slope downward (i.e., have a negative slope) and that the supply curve must slope upward (i.e., have a positive slope). By adding an equilibrium condition that the quantity supplied equals the quantity demanded, that is,

$$Q_d = Q_s \tag{2-8}$$

the economic model, consisting of equations (2-6), (2-7), and (2-8), is complete.

By equating supply and demand, the equilibrium price ($P_e$) can be determined. Substituting equations (2-6) and (2-7) for $Q_d$ and $Q_s$ in (2-8) yields

$$a + bP = c + dP$$

Solving for $P$, the equilibrium price is

$$P_e = \frac{c - a}{b - d}$$

The equilibrium quantity is found by substituting the equilibrium price into either the supply or demand function and solving for quantity. Using the demand function, the equilibrium quantity is

$$Q_e = a + b\left(\frac{c - a}{b - d}\right)$$

If the values of the parameters $a$, $b$, $c$, and $d$ are known, the actual values of $P_e$ and $Q_e$ can easily be calculated. In chapter 4, statistical methods used to estimate the numerical values of these parameters are reviewed.

Models of this type are used extensively in the study of economics. Indeed, one of the strengths of the discipline is that important and powerful results can be derived from models that are quite simple. To be sure, some economic models in advanced books are very complex. However, a thorough grasp of many of the key principles of economics can be obtained by applying rather basic concepts and models to business and social problems. Clearly, an important reason for studying managerial economics is to develop a set of economic models that can be used to analyze the many resource allocation problems faced by managers.

### Example Determining Equilibrium Price and Quantity

Suppose parameters of the demand and supply equations have been estimated and that the equations are

$$Q_d = 14 - 2P$$

and

$$Q_s = 2 + 4P$$

Determine the equilibrium price and quantity.

**Solution** Note that the slopes of these functions meet the constraints specified in the chapter. That is, the demand function slopes downward and the supply curve slopes upward. Setting the demand function equal to the supply function

$$14 - 2P = 2 + 4P$$

and solving for the price yields

$$P = 2$$

Substituting $P_e = 2$ into the demand function yields the equilibrium quantity

$$Q_e = 14 - 2(2) = 10$$

Thus the {price, quantity} combination $\{P_e = 2, Q_e = 10\}$ results in equilibrium in this market. This combination corresponds to $\{P_e, Q_e\}$ in Figure 2.2.

### Key Concepts

- An economic model typically consists of several functional relationships, conditions, or constraints on one or all of these functions, and one or more equilibrium conditions.
- Generally, economic models are used to demonstrate an economic principle, to explain an economic phenomenon, or to predict the economic implications of some change affecting one or more of the functional relationships.

## PROBABILITY AND PROBABILITY DISTRIBUTIONS

Managers are often faced with making decisions that have the potential for a variety of outcomes. An *outcome* is a possible result of some action. For example, flipping a coin will result in one of two outcomes—heads or tails. A management decision to introduce a new product could result in a range of outcomes varying from wide consumer acceptance to no interest whatsoever. The quality of any decision will be enhanced by identifying the possible outcomes of the decision and then estimating the relative chances of each occurring. This listing of outcomes and the chance of each occurring is a *probability distribution* that can be evaluated using quantitative techniques. Often the chance of an outcome occurring will depend on the state of nature that prevails. The term *state of nature* refers to conditions in the business environment that will influence the outcome of a decision but that are not subject to control by the decision maker. For example, the condition of the general economy is a state of nature that will affect the outcomes of most business decisions.

Probability and probability distributions are an important part of the manager's tool kit. In this section, methods are developed to set up the probability distribution and then evaluate measures of central tendency and dispersion for that distribution. In chapter 14, these principles are extended to analyze the concept of risk in decision making.

### Probability

The *probability* of an event is the relative frequency of its occurrence in a large number of repeated trials. For example, in repeated tossing of a coin, a head will appear about one-half the time. That is, the relative frequency or probability of a head occurring is 0.50. In rolling a die, there are six possible outcomes, 1, 2, 3, 4, 5, 6. The relative frequency of any one of these outcomes is 1/6, implying a probability of 0.167 for each outcome.

The probability of some events is known or can be computed with certainty. For example, the probabilities of most outcomes associated with rolling dice, drawing playing cards from a deck, and tossing coins are easily determined using standard principles of probability. Other probabilities are not easily determined mathematically but can be determined by repeating the process many times and observing how many times particular outcomes occur. For instance, there may be no mathematical way to estimate the probability of winning a particular game. But by playing the game many times and noting the number of times it is won, the probability of winning can be determined.

In other cases there is no accurate way to estimate probabilities except by using judgment. The weather forecaster's statement that "the probability of measurable precipitation tomorrow is 0.40" is based on an evaluation of how prevailing breezes, radar maps, and upper-air charts are associated with given weather conditions. Assessing the probability of a recession next year or a significant change in consumer preferences falls into the same category. Based on an analysis of current economic conditions, surveys of business capital-spending plans, and other information, a judgment can be made as to that probability. This judgment is necessarily subjective and may differ significantly among analysts, but predictions of this type are made daily by business managers, economists, and other decision makers.

## Probability Distributions

For a decision or an experiment of some type with several possible outcomes, the probability of the $i$th outcome occurring is indicated by $P_i$, where

$$0 \le P_i \le 1 \quad i = 1, 2, \ldots, n \tag{2-9}$$

That is, the probability must take on a value in the range 0 to 1. Negative probabilities and those in excess of unity have no meaning. For example, if an event had a probability of 1.25, that would mean that it would occur more than 100 percent of the time!

Furthermore, the sum of the probabilities of all possible outcomes or events of an experiment must equal 1. That is,

$$\sum_{i=1}^{n} P_i = 1 \tag{2-10}$$

This means that when the experiment is conducted (i.e., the dice rolled, the coin tossed, or the investment made), one of the outcomes must occur. It then follows that the probability that an event will *not* occur is 1 minus the probability that it will occur. For example, if the probability of a 6 occurring when rolling a die is 1/6, this implies that the probability of not rolling a six is 1 − 1/6 = 5/6. This corollary is useful because it is sometimes easier to find the probability of an event not occurring than the probability that the event will occur.

A listing of each outcome of an experiment and its probability defines a probability distribution. For example, the probabilities of tossing 0, 1, 2, or 3 heads in three tosses of a coin are shown in Table 2.2. In this example, $X_i$ is the number of heads observed in each experiment of three tosses. Note that each probability is between 0 and 1 and that the sum of the probabilities is equal to 1.

**TABLE 2.2 Probabilities Associated with Tossing a Coin Three Times**

| *Number of Heads* ($X_i$) | *Probability* ($P_i$) |
|---|---|
| 0 | 0.125 |
| 1 | 0.375 |
| 2 | 0.375 |
| 3 | 0.125 |

## Statistics of a Probability Distribution

For any probability distribution, there is a set of statistics or measures that describes or provides summary information about the distribution. The most important of these are measures of central tendency and dispersion. These have many applications in business, science, and engineering. In managerial economics they can be used to evaluate and compare the returns associated with alternative strategies and thus help to make sound managerial decisions.

*Expected Value:* The first of these statistics is the expected value or mean of a probability distribution. Effectively, it is a weighted average of the outcomes using the probabilities of those outcomes as weights. The expected value is a measure of *central tendency* because in repeated trials of most experiments, the values of the outcomes tend to be concentrated around this statistic. For example, the expected value of the distribution for rolling two dice is 7. The range of outcomes is 2 through 12, but the outcomes 6 through 8 occur much more often than the outcomes 2, 3, 11, or 12. This is because there are more combinations of the faces of the dice that yield 6, 7, or 8 than combinations that yield 2, 3, 11, or 12. There is only one combination (1, 1) that yields 2, but there are six combinations (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1) that yield 7.

The expected value ($\mu$) of any probability distribution is computed by multiplying each outcome by its respective probability and then summing the products. Thus

$$\mu = P_1 X_1 + P_2 X_2 + \ldots + P_n X_n = \sum_{i=1}^{n} P_i X_i \tag{2-11}$$

Recall the probability distribution for the number of heads observed in three tosses of a coin.

| *Heads* | *Probability* |
|---|---|
| 0 | 0.125 |
| 1 | 0.375 |
| 2 | 0.375 |
| 3 | 0.125 |

The expected value of this distribution is computed to be

$$\mu = 0.125(0) + 0.375(1) + 0.375(2) + 0.125(3) = 1.50$$

This means that if this experiment were repeated many times and the number of heads observed for each trial recorded, the average number of heads would be 1.50. Obviously, it is not possible to toss 1.5 heads on any trial. The expected value is not a particular outcome but is merely an average of the outcomes for a large number of repetitions of the experiment.

Understanding this principle and applying it when making managerial plans and decisions can be very profitable. The financial success of gambling casinos in Las Vegas and Atlantic City is evidence of the truth of this statement. The managers of these firms use the principles of probability to structure the gambling games played in their casinos. The payoffs to the players are always set so that the house wins more than an equal share of the amounts wagered. The return on a gambling game is often referred to as the expected payoff. It is computed by multiplying the probability of winning by the amount to be won and adding the product of the probability of losing and the amount to be lost. Consider a slot machine that costs \$1 per play. The machine is programmed so that on average it returns \$8 once in every 10 plays for a net payoff of \$7 (i.e., the \$8 gross return less the \$1 cost of playing the game). The probability distribution is as follows:

| *State of Nature* | *Outcome* | *Probability* |
|---|---|---|
| Win | \$ 7 | 0.10 |
| Lose | −1 | 0.90 |

The expected payoff to the player is \$−0.20, that is,

$$\mu = 0.10(7) + 0.9(-1) = -0.20$$

A game is said to be fair if the expected return or payoff is zero. Clearly, the previous example is not a fair game; the player expects to lose \$0.20 (equivalently, the casino expects to win \$0.20) every time the slot machine is played. Gambling games in casinos are not fair in a statistical sense because the expected values are negative for the player.

*Standard Deviation:* The second statistic of interest measures the dispersion of possible outcomes around the expected value. This measure is called the standard deviation ($\sigma$) and is given by the equation

$$\sigma = \sqrt{\Sigma P_i(X_i - \mu)^2} \tag{2-12}$$

As just indicated, the standard deviation measures the variability of outcomes around the expected value of the distribution. If there is no variation in the outcomes, that is, if all $X_i = \mu$, then $\sigma = 0$. As the amount of variation about the expected value increases, the value of $\sigma$ also increases.

The computation of $\sigma$ for the probability distribution for tossing three coins is as follows:

$$\begin{aligned}\sigma &= \sqrt{0.125(0 - 1.5)^2 + 0.375(1 - 1.5)^2 + 0.375(2 - 1.5)^2 + 0.125(3 - 1.5)^2}\\ &= 0.87\end{aligned}$$

Because the standard deviation provides information about the dispersion of the individual values or outcomes around the expected value, this measure of variation can be of use in decision making. Suppose that a manager was considering two investment alternatives, $A$ and $B$. The probability distributions for each are shown in Figure 2.3. Note that the expected value of profit (i.e., \$100) is the same for both, but there is a much greater range of outcomes for $B$ than $A$. This implies that $\sigma_B$ is greater than $\sigma_A$

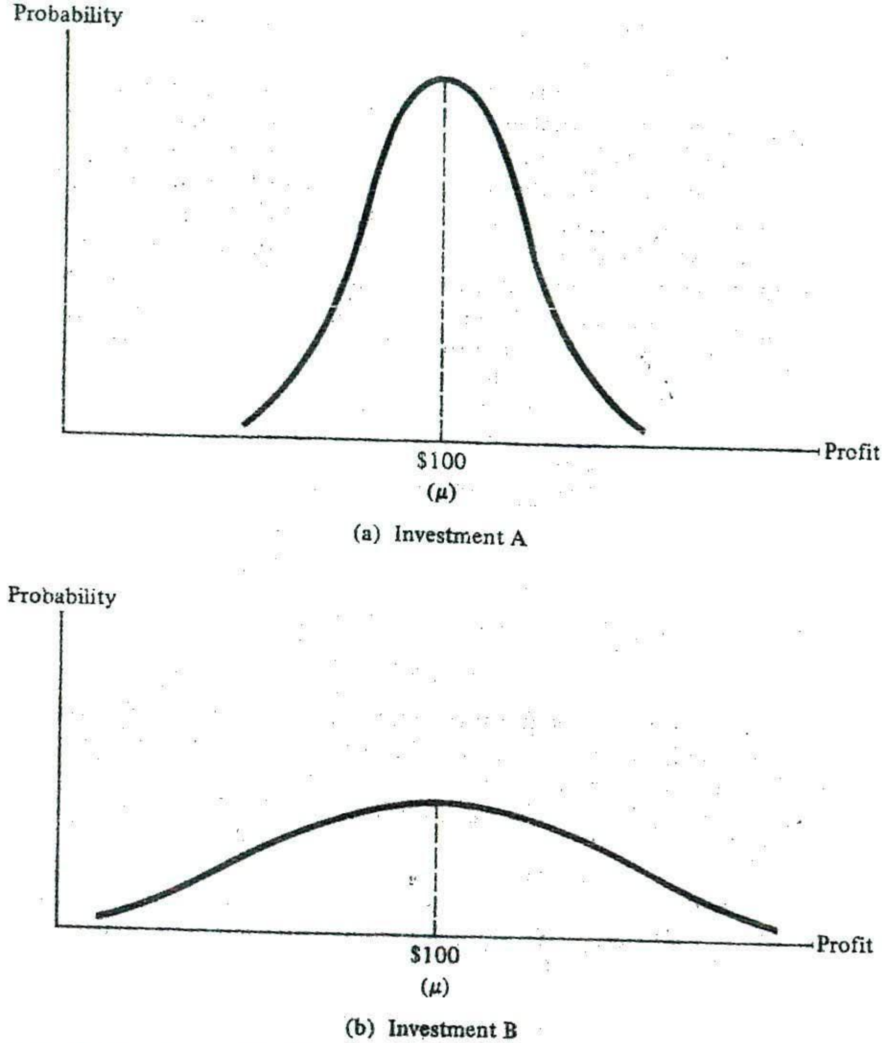



FIGURE 2.3 Probability Distributions for the Profits Associated with Two Alternative Investments

(i.e., $B$ is a riskier investment than $A$). Most decision makers will pick $A$, although there may be others who would select $B$ because of differences in their preference for risk. The nature of these preferences is discussed in more detail in chapter 14.

*Coefficient of Variation:* The third statistic is the coefficient of variation, which relates the variation of outcomes to the mean. It is defined as the ratio of the standard deviation to the expected value:

$$v = \frac{\sigma}{\mu} \tag{2-13}$$

This statistic provides a useful way to compare the variation to the expected value of a probability distribution. It measures variation per unit of expected value. The use of the coefficient of variation can be illustrated by considering two probability distributions, $C$ and $D$, having the following statistics:

$$\mu_C = 100 \qquad \sigma_C = 50$$
$$\mu_D = 50 \qquad \sigma_D = 30$$

Both the mean and standard deviation for distribution $C$ are greater than for distribution $D$, but the coefficient of variation is less than for $C$ than for $D$. That is,

$$\upsilon_C \frac{50}{100} = 0.5 < \upsilon_D = \frac{30}{50} = 0.6$$

This means that distribution $C$ has less dispersion or risk relative to its mean than does distribution $D$. Thus, not only does the absolute amount of risk ($\sigma$) have to be estimated and evaluated, but the level of that risk relative to the expected return on an investment also should be considered.

**Key Concepts**

- A probability distribution lists the possible outcomes of an experiment and the probability associated with each outcome.
- The probability of any outcome must be in the range $0 \le P(X_i) \le 1.0$, and the sum of the probabilities for all outcomes must equal 1.
- The expected value of a probability distribution ($\mu = \Sigma P_i X_i$) is a measure of the average outcome from repeated trials of an experiment.
- The standard deviation $[\sigma = \sqrt{\Sigma P_i (X_i - \mu)^2}\,]$ measures the dispersion of outcomes around the expected value and is used to assess risk.
- The coefficient of variation ($\upsilon = \sigma/\mu$) is a measure of variation per unit of expected value.

## SUMMARY

A functional relationship such as $y = f(x_1, x_2, \ldots, x_n)$ means that the value of the dependent variable $y$ depends on the values of the independent variables $x_1, x_2, \ldots, x_n$ in a systematic way and that there is a unique value of $y$ for any set of values for those independent variables. Total, average, and marginal relationships are fundamental to understanding and using economics. The marginal function is the slope of the total function. The intersection of the marginal and average functions occurs at either the minimum or the maximum point of the average function.

An economic model generally consists of several related functions and certain conditions related to these functions. The model can be used to explain economic phenomena and to predict the effects of changes in one or more variables or functions. Often, simple models can be used to explain complex relationships and to derive powerful results. The effectiveness of managerial decision making can be enhanced significantly by using economic models.

An outcome is a possible result of some action. A state of nature is a condition that affects the outcome but cannot be controlled by the decision maker. The probability of an event happening is its relative frequency of occurrence. The probability of an outcome must be in the range $0 \leq P(X_i) \leq 1.0$ and the sum of all probabilities must equal one.

The combination of all possible outcomes and their probabilities comprises a probability distribution. For any such distribution there is a set of statistics. The expected value or mean of a probability distribution measures central tendency. The standard deviation measures the dispersion of the outcomes, and the coefficient of variation provides information on the amount of dispersion relative to the central tendency of the distribution. The use of probability distributions and their associated statistics has wide applicability in managerial decision making.

## Discussion Questions

2-1. Explain the concept of an economic model. Why do economists and managers use such models as part of the decision-making process?

2-2. Economists are not the only scientists to use models in their work. Describe how other disciplines use models or similar abstractions from reality in their work.

2-3. Explain the relationship among the total, average, and marginal functions. Intuitively explain why any intersection of the average and the marginal functions will occur at a maximum or a minimum point on the average function.

2-4. The president of a major firm (who has had no training in managerial economics) complains during a board meeting that quantitative techniques are of little value because they are always subject to error and therefore should not be part of the decision-making process. If the chairman asked for your opinion, what would be your response?

2-5. Explain how principles of probability are used to set the rules and payoffs for various gambling games in casinos.

2-6. Is the statement "$y$ is a function of $x$" equivalent to saying that "$y$ is caused by $x$"? In this context, critically evaluate the following statement: "The incidence of lung cancer is significantly higher for heavy smokers than for nonsmokers. Therefore, smoking causes cancer."

2-7. Explain how an insurance company could use its historical experience on deaths and accidents to set its insurance rates.

2-8. The General Mills Company has several brands of breakfast cereal on the market. Explain how you could use information on historic sales of these products to develop a probability distribution for the sales of a new brand of cereal.

## Problems

2-1. Given the following supply and demand equations

$$Q_D = 100 - 5P$$
$$Q_S = 10 + 5P$$

where $Q_D$ and $Q_S$ represent the quantities demanded and supplied, respectively, and $P$ is the price.

a. Determine the equilibrium price and quantity.

b. If the government sets a minimum price of \$10 per unit, how many units would be supplied and how many would be demanded? How could the government maintain this minimum price?

c. If the government sets a maximum price of \$5 per unit, how many units would be supplied and how many would be demanded?

d. If demand increases to

$$Q'_D = 200 - 5P$$

determine the new equilibrium price and quantity.

2-2. Given the following demand function

$$Q = 20 - 0.10P$$

where $P$ = price and $Q$ = rate of output, complete the following table. (Note that total revenue is equal to price times quantity.)

| *Quantity* | *Price* | *Total Revenue* | *Average Revenue* | *Marginal Revenue* |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |

2-3. Given the total cost function

$$TC = 150Q - 3Q^2 + 0.25Q^3$$

complete the following table by computing the total, average, and marginal costs associated with each quantity indicated.

| *Quantity* | *Total Cost* | *Average Cost* | *Marginal Cost* |
|---|---|---|---|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |

2-4. Using the data developed in Problems 2-2 and 2-3,

a. Plot the total, average, and marginal functions for both revenue and cost. [*Note:* Use a two-part graph similar to Figure 2.1 that places the total function in the upper part and the average and marginal functions in the lower half. Also, it is conventional to plot the marginal values at the midpoint between the quantity values to which they relate. For example, the marginal revenue associated with going from two units of output to three units is $150. This value should be plotted midway between the two- and three-unit marks on the quantity (i.e., horizontal) axis.]

b. What is true of the marginal revenue function at that level of output where total revenue is at a maximum?

c. Determine total profit by subtracting total cost from total revenue at each rate of output. Plot this function in the upper part of your diagram.

d. By comparing the marginal revenue and marginal cost curves and relating them to the profit curve, can you think of a rule the firm might use to determine the output rate that would maximize profit?

2-5. Determine the average function for each of the following total functions:

a. Total revenue $= 100Q - Q^2$

b. Total cost $= 1{,}000 + 10Q + 0.01Q^2$

c. Total profit $= 50Q - 0.1Q^2 - 1{,}000$

2-6. Given the following total revenue ($TR$) and total cost ($TC$) equations, determine the output rate that would result in a breakeven (i.e., zero profit) situation for the firm.

$$TR = 51Q - Q^2$$
$$TC = 625 + Q$$

2-7. Do each of the following distributions meet the requirements for a probability distribution? Why or why not?

a.

| $X_i$ | $P(X_i)$ |
|---|---|
| 10 | 0.10 |
| −20 | 0.20 |
| −30 | 0.30 |
| −50 | 0.40 |

b.

| $Y_i$ | $P(Y_i)$ |
|---|---|
| 5 | −0.20 |
| 10 | 0.40 |
| 20 | 0.40 |
| 40 | 0.40 |

c.

| $Z_i$ | $P(Z_i)$ |
|---|---|
| −2 | 0.20 |
| −6 | 0.30 |
| −8 | 0.40 |
| −10 | 0.20 |

2-8. Given the following probability distribution

| $X_i$ | $P(X_i)$ |
|---|---|
| 2 | 0.10 |
| 4 | 0.20 |
| 6 | 0.30 |
| 8 | 0.30 |
| 10 | 0.10 |

Compute the expected value, standard deviation, and the coefficient of variation.

2-9. A firm is contemplating building a new factory that will have three possible levels of profit, depending on business conditions. The possible levels of profit and the probability of each occurring are

| *Profit* | *Probability* |
|---|---|
| −1,000 | 0.20 |
| 5,000 | 0.30 |
| 10,000 | 0.50 |

Determine the expected value, standard deviation, and the coefficient of variation for this probability distribution.

2-10. An advertising program has three possible outcomes: excellent success (i.e., 1,000 units sold); good success (i.e., 500 units sold); and little success (i.e., 100 units sold). The probability of excellent success is 0.4 and the probability of little success is 0.2. Find the mean, standard deviation, and coefficient of variation for this probability distribution.

## Problems Requiring Calculus

2-11. A firm's total revenue ($TR$) and total cost ($TC$) functions are

$$TR = 110Q - 5Q^2$$
$$TC = 10Q - Q^2 + 0.33Q^3$$

Determine:

a. Equations for marginal revenue and average revenue.
b. Equations for marginal cost and average cost.
c. The output rate that maximizes total revenue.

A-62974

d. The output rate that minimizes average cost.
e. The output rate that maximizes profit.

2-12. Given the following function that relates total revenue ($TR$) to output ($Q$),

$$TR = 20Q - 2Q^2$$

determine

a. That rate of output that results in maximum total revenue.
b. The marginal revenue function.
c. The rate of output for which marginal revenue is zero.
d. If there is any connection between your answers to parts (a) and (c)? Explain.

2-13. Given the total cost function

$$TC = 100Q - Q^2 + 0.3Q^3$$

where $Q$ = rate of output and $TC$ = total cost, determine

a. The marginal and average cost functions.
b. The rate of output that results in minimum average cost.

2-14. Given the firm's demand function

$$Q = 55 - 0.5P$$

(where $P$ = price and $Q$ = rate of output), and the total cost function

$$TC = 20 + Q + 0.2Q^2$$

where $TC$ = total cost, determine

a. The total revenue function for the firm. (*Hint:* To find the total revenue function, solve the demand function for $P$ and then multiply both sides of the equation by $Q$.)
b. The marginal revenue and marginal cost functions and find the rate of output for which marginal revenue equals marginal cost.
c. An equation for profit by subtracting the total cost function from the total revenue function. Find the level of output that maximizes total profit. Compare your answer to that obtained in part (b). Is there any correspondence between these answers?

2-15 The profit function for a firm selling two products is

$$\pi = 50Q_1 - Q_1^2 + 100Q_2 - 4Q_2^2$$

where $Q_1$ and $Q_2$ represent output rates of products 1 and 2, respectively. Determine the profit maximizing output rates for the two products and the profit associated with these output rates.

# Appendix

## *Calculus and Managerial Economics*

Recall that many of the decisions facing managers fall into the category of optimization problems. For example, decisions relating to maximizing profit or minimizing cost clearly involve optimization. Often such problems can be solved graphically or by using algebra. In other cases, however, the solution requires the use of calculus. More importantly, in many cases calculus can be used to solve such problems more easily and with greater insight into the economic principles underlying the solution.

Consider a firm whose total revenues from sales are given by the function

$$TR = 20Q - Q^2$$

where $Q$ represents the rate of output. Assume that the total cost of producing any rate of output is given by the equation

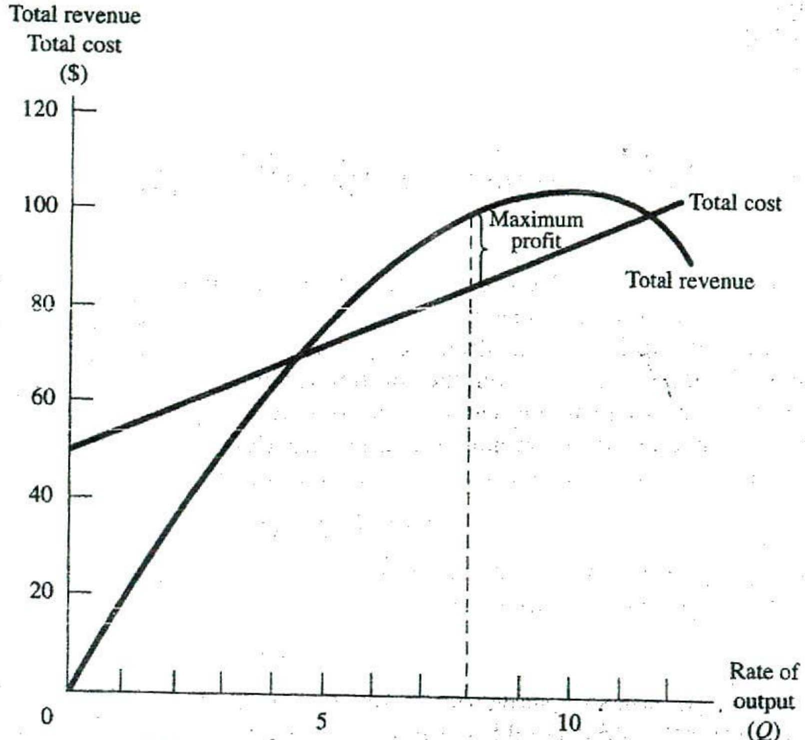
$$TC = 50 + 4Q$$

Given the firm's objective of maximizing profits (i.e., total revenue minus total cost), how much output should be produced? One approach would be to graph both the total revenue and total cost functions. The vertical distance between the two functions is profit. By identifying that point where this vertical distance is the largest, the profit-maximizing output is found. This is shown in Figure 2A.1, which shows profit is maximized by producing eight units of output ($Q = 8$).

An alternative approach is to develop a table showing revenue, cost, and profit at each rate of output. Such data are provided in Table 2A.1. The tabular method has the advantage of being somewhat more precise. That is, at an output rate of 8, total revenue is 96, total cost is 82, and profit is 14. From the graph in Figure 2A.1, these values can only be approximated.

The problem is that neither method is very efficient. What if the profit-maximizing output level is 8,000 or 8,000,000 instead of 8? The answer could be found using either approach, but finding that answer could take considerable time. What if there had been two outputs ($Q_1$ and $Q_2$) and three inputs (land, labor, and capital)? In this case, there would be no practical way to determine profit-maximizing output rates for $Q_1$ and $Q_2$ using these methods.

A more powerful technique is needed so that the solution process can be both precise and straightforward. Elementary calculus is easily adapted to optimization problems in economics. Indeed, a few basic principles can be used in many different kinds of problems. The profit-maximization problem just outlined could have been solved quickly using the most elementary calculus. In the following pages, some basic principles of calculus are outlined and their application to economic problems is demonstrated.




[redacted]

[redacted]

| $Q$ | $TR$ | $TC$ | Profit ($TR - TC$) |
|---|---|---|---|
| 0 | 0 | 50 | −50 |
| 1 | 19 | 54 | −35 |
| 2 | 36 | 58 | −22 |
| 3 | 51 | 62 | −11 |
| 4 | 64 | 66 | −2 |
| 5 | 75 | 70 | 5 |
| 6 | 84 | 74 | 10 |
| 7 | 91 | 78 | 13 |
| 8 | 96 | 82 | 14 |
| 9 | 99 | 86 | 13 |
| 10 | 100 | 90 | 10 |
| 11 | 99 | 94 | 5 |

## THE DERIVATIVE OF A FUNCTION

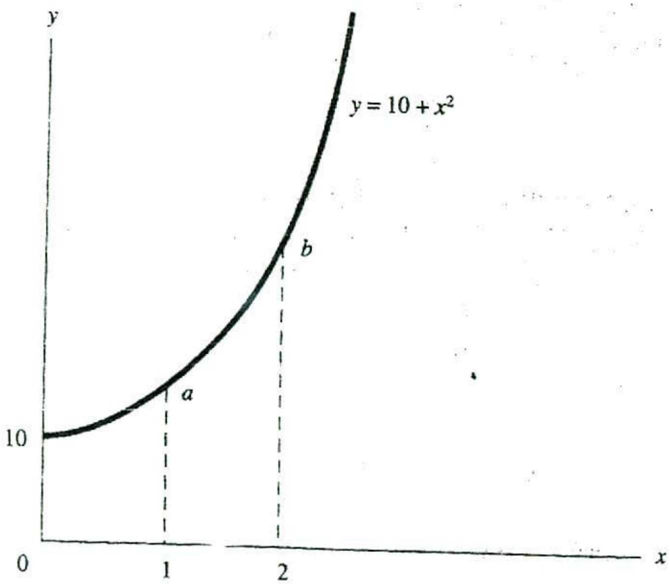
From algebra recall that for the function $y = f(x)$, the slope of that function is th change in $y$ (denoted by $\Delta y$) divided by a change in $x$ (i.e., $\Delta x$). The slope sometimes i referred to as the *rise* (the change in the variable measured on the vertical axis) ove the *run* (the change in the variable measured on the horizontal axis). The slope is *posi tive* if the curve slopes upward from left to right and *negative* if the function slope: downward from left to right. A horizontal line has a *zero slope,* and a vertical line is saic to have an *infinite slope*. For a positive change in $x$ (i.e., $\Delta x > 0$), a positive slope implies that $\Delta y$ is positive, and a negative slope implies that $\Delta y$ is negative.

The function $y = 10 + x^2$ is graphed in Figure 2A.2. To determine the average slope of this function over the range $x = 1$ to $x = 2$, first find the corresponding $y$ values. If $x_1 = 1$, then $y_1 = 11$, and if $x_2 = 2$, then $y_2 = 14$. Then the slope is found by using the formula

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{14 - 11}{2 - 1} = 3$$

In reality, this method determines the slope of a straight line through the points *a* and *b* in Figure 2A.2. Thus, it is only a rough approximation of the slope of the function $y = 10 + x^2$, which actually changes at every point on that function. By making the interval smaller, a better estimate of the slope is determined. For example, consider the slope over the interval $x = 1$ to $x = 1.1$. If $x_1 = 1$, then $y_1 = 11$, and $x_2 = 1.1$ implies that $y_2 = 11.21$. Thus, the slope is

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{11.21 - 11}{1.1 - 1} = \frac{0.21}{0.10} = 2.1$$




[illegible]

By using calculus, the exact slope at any point on the function can be determined.

The first derivative of a function (denoted as $dy/dx$) is simply the slope of a function when the interval along the horizontal axis is infinitesimally small. Technically, the derivative is the limit of the ratio $\Delta y/\Delta x$ as $\Delta x$ approaches zero, that is,

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

The derivative of $y = f(x)$ is sometimes written as $f'(x)$. The calculus term $dy/dx$ is analogous to $\Delta y/\Delta x$, but $dy/dx$ is the precise slope at a point, whereas $\Delta y/\Delta x$ is the average slope over an interval of the function. The derivative can be thought of as the slope of a straight line drawn tangent to the function at that point. For example, the slope of the function $y = 10 + x^2$ at point $a$ in Fugure 2A.2 is the slope of th;e straight line drawn tangent to that function at $a$.

What is the significance of this concept for managerial economics? Recall from the discussion of total and marginal relationships that the marginal function is simply the slope of the total function. Calculus offers an easy way to find the marginal function by taking the first derivative of the total function. Calculus also offers a set of rules for using these derivatives to make optimizing decisions such as minimizing cost or maximizing profit.

Standard calculus texts present numerous formulas for the derivative of various functions. In the hands of the skilled mathematician, these formulas allow the derivative of virtually any function to be found. However, only a few of these rules are necessary to solve most of the relevant problems in managerial economics. In this section each of these basic rules is explained and its use demonstrated.

### The Derivative of a Constant

The derivative of any constant is zero. When plotted, the equation of a constant (such as $y = 5$) is a horizontal line. For any $\Delta x$, the change in $y$ is always zero. Thus for any equation $y = a$, where $a$ is a constant,

$$\frac{dy}{dx} = 0 \qquad \text{(2A-1)}$$

### The Derivative of a Constant Times a Function

The derivative of a constant times a function is that constant times the derivative of the function. Thus the derivative of $y = af(x)$, where $a$ is constant, is

$$\frac{dy}{dx} = af'(x) \qquad \text{(2A-2)}$$

For example, if $y = 3x$, the derivative is

$$\frac{dy}{dx} = 3f'(x) = 3(1) = 3$$

### The Derivative of a Power Function

For the general power function of the form $y = ax^b$, the derivative is

$$\frac{dy}{dx} = bax^{b-1} \qquad \text{(2A-3)}$$

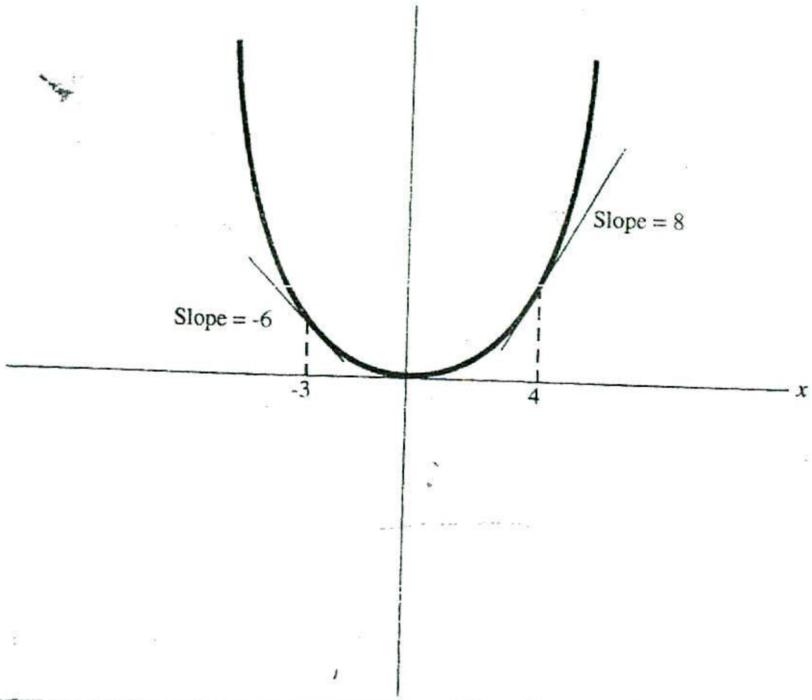



FIGURE 2A.3 Graph of Function $y = x^2$

The function $y = x^2$ is a specific case of a power function where $a = 1$ and $b = 2$. Hence the derivative of this function is

$$\frac{dy}{dx} = 2x^{2-1} = 2x$$

The interpretation of this derivative is that the slope of the function $y = x^2$ at any point $x$ is $2x$. For example, the slope at $x = 4$ would be found by substituting $x = 4$ into the derivative. That is,

$$\left(\frac{dy}{dx}\right)_{x=4} = 2(4) = 8$$

Thus when $x = 4$, the change in $y$ is 8 times a small change in $x$.

The function $y = x^2$ is shown in Figure 2A.3. Note that the slope changes continually. The slope at $x = 4$ is 8. As $x$ increases, the slope becomes steeper. For negative values of $x$, the slope is negative. For example, if $x = -3$, the slope is $-6$.

## The Derivative of a Sum or Difference

The derivative of a function that is a sum or difference of several terms is the sum (or difference) of the derivatives of each of the terms. That is, if $y = f(x) + g(x)$, then

$$\frac{dy}{dx} = f'(x) + g'(x) \tag{2A-4}$$

For example, the derivative of the function

$$y = 10 + 5x + 6x^2$$

is equal to the sum of the derivatives of each of the three terms on the right-hand side. Note that the rules for the derivative of a constant, a constant times a function, and a power function must be used. Thus

$$\frac{dy}{dx} = 0 + (1)5x^0 + (2)6x^1 = 5 + 12x$$

Consider another function: $y = 2x^3 - 6x^{-2} - 4x + 10$

The derivative is $\dfrac{dy}{dx} = (3)2x^2 - (-2)6x^{-3} - 4 - 0$

$$6x^2 + 12x^{-3} - 4$$

## The Derivative of the Product of Two Functions

Given a function of the form $y = f(x)g(x)$, the derivative $dy/dx$ is given by

$$\frac{dy}{dx} = f'(x)g(x) + f(x)g'(x) \tag{2A-5}$$

That is, the derivative of the product of two functions is the derivative of the first function times the second function plus the first function times the derivative of the second.

For example, the derivative of the function

$$y = (x^2 - 4)(x^3 + 2x + 2)$$

is

$$\frac{dy}{dx} = (2x)(x^3 + 2x + 2) + (x^2 - 4)(3x^2 + 2) = 5x^4 - 6x^2 + 4x - 8$$

## The Derivative of a Quotient of Two Functions

For a function of the form $y = f(x)/g(x)$, the derivative is

$$\frac{dy}{dx} = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2} \tag{2A-6}$$

Given the function

$$y = \frac{x^2 - 3x}{x^2}$$

the derivative would be

$$\frac{dy}{dx} = \frac{x^2(2x - 3) - (x^2 - 3x)(2x)}{(x^2)^2} = \frac{3x^2}{x^4} = \frac{3}{x^2}$$

## The Derivative of a Function of a Function

The function

$$y = (2x + 5)^3$$

is really two functions combined. That is, by writing

$$u = f(x)$$

where

$$u = 2x + 5$$

and

$$y = g(u)$$

where

$$y = u^3$$

it is seen that $y$ is a function of a function. That is,

$$y = g(u) = g[f(x)]$$

This derivative is the derivative of $y$ with respect to $u$ and multiplied by the derivative of $u$ with respect to $x$, or

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Now, using the rule for the derivative of a power function yields

$$\frac{dy}{du} = 3u^2 = 3(2x + 5)^2$$

and

$$\frac{du}{dx} = 2$$

so

$$\frac{dy}{dx} = [3(2x + 5)^2] \cdot 2 = 6(2x + 5)^2$$

Consider another example:

$$y = \frac{1}{\sqrt{x^5 + 2x + 6}}$$

which can be rewritten as

$$y = (x^5 + 2x + 6)^{-1/2}$$

In this case,

$$y = u^{-1/2}$$

and

$$u = x^5 + 2x + 6$$

so the solution is

$$\frac{dy}{du} = -\frac{1}{2}u^{-(1/2)-1} = -\frac{1}{2}u^{-3/2}$$

and

$$\frac{du}{dx} = 5x^4 + 2$$

Substituting $x^5 + 2x + 6$ for $u$ and multiplying the two derivatives just given yields

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = -\frac{1}{2}(x^5 + 2x + 6)^{-3/2}(5x^4 + 2)$$

These seven rules of differentiation are sufficient to determine the derivatives of all the functions used in this book. However, sometimes two or more of the rules must be used at the same time.

## Example Finding the Marginal Function

Given a total revenue function

$$TR = 50Q - 0.5Q^2$$

and a total cost function

$$TC = 2{,}000 + 200Q - 0.2Q^2 + 0.001Q^3$$

find the marginal revenue and marginal cost functions.

**Solution** Recall that a marginal function is simply the slope of the corresponding total function. For example, marginal revenue is the slope of total revenue. Thus, by finding the derivative of the total revenue function, the marginal revenue function will be obtained:

$$MR = \frac{d(TR)}{dQ} = 50 - 2(0.5)Q^{2-1}$$
$$= 50 - Q$$

Similarly, the marginal cost function will be found by taking the first derivative of the total cost function:

$$MC = \frac{d(TC)}{dQ} = 0 + 200 - 2(0.2)Q^{2-1} + 3(0.001)Q^{3-1}$$
$$= 200 - 0.4Q + 0.003Q^2$$

### Key Concepts

- The slope of a function $y = f(x)$ is the change in $y$ (i.e., $\Delta y$) divided by the corresponding change in $x$ (i.e., $\Delta x$).
- For a function $y = f(x)$, the derivative, written as $dy/dx$ or $f'(x)$, is the slope of the function at a particular point on the function. Equivalently, the derivative is the slope of a straight line drawn tangent to the function at that point.
- By using one or more of the seven formulas outlined in this appendix, the derivative of most functions encountered in managerial economics can be found.

## HIGHER-ORDER DERIVATIVES

The derivative of a function sometimes is called the *first derivative* to indicate that there are higher-order derivatives. The *second derivative* of a function is simply the first derivative of the first derivative; it is written $d^2y/dx^2$ or $f''$. In the context of economics, the first derivative of a total function is the marginal function. The second derivative of the total function is the slope of the first derivative or the rate at which the marginal function is changing.

Higher-order derivatives are easy to find. One simply keeps taking the first derivative again. Given the function

$$y = 10x^3 + 3x^2 - 5x + 6$$

the first derivative is

$$\frac{dy}{dx} = 30x^2 + 6x - 5$$

and the higher-order derivatives are

$$\textit{second:}\quad \frac{d^2y}{dx^2} = 60x + 6$$

$$\textit{third:}\quad \frac{d^3y}{dx^3} = 60$$

$$\textit{fourth:}\quad \frac{d^4y}{dx^4} = 0$$

The second derivative has an important application in finding the maximum and/or minimum of a function. This concept is explained in the following section.

## CALCULUS AND OPTIMIZATION

Recall from the discussion of total and marginal relationships that if the marginal function is positive, the total function must be increasing; if the marginal function is negative, the total function must be decreasing. It was shown that if the marginal function is zero, then the total function must be at either a maximum or a minimum. In Figure 2A.4, a total function and its associated marginal function are shown. At point *a*, which corresponds to $x = x_1$, the total function is at a maximum and the marginal function is zero. At point *b*, corresponding to $x = x_2$, the total function is minimized and the marginal function is again zero. Thus the marginal curve is zero at both $x_1$ and $x_2$. However, note the difference in the slope of the marginal function at these points. At $x_1$, the marginal curve has a negative slope, whereas at $x_2$, the slope is positive.

Because the total function is at a maximum or a minimum (i.e., an extremum) when its slope is zero, one way to find the value of $x$ that results in a maximum or a minimum is to set the first derivative of the total function equal to zero and solve for $x$. This is a better approach than the trial-and-error method used earlier. In that example, a total revenue function,

$$TR = 20Q - Q^2$$

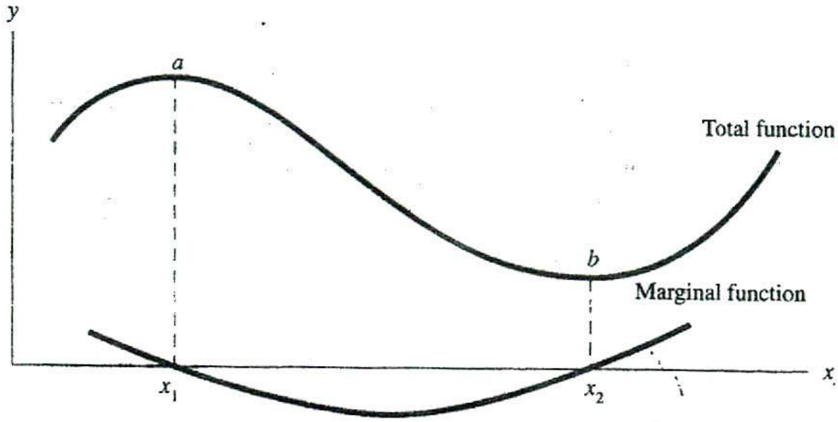



**FIGURE 2A.4 Relationship of Total and Marginal Functions at Minimum and Maximum Points**

and a total cost function,

$$TC = 50 + 4Q$$

were given. The problem was to find the rate of output, $Q$, that maximized profit.

The total profit function ($\pi$) is found by subtracting total cost from total revenue:

$$\begin{aligned}\pi &= TR - TC \\ &= 20Q - Q^2 - 50 - 4Q \\ &= -Q^2 + 16Q - 50\end{aligned}$$

The profit function will have a slope of zero where that function is at a maximum and also at its minimum point. To find the profit-maximizing output, take the first derivative of the profit function with respect to output, set that derivative equal to zero, and solve for output. That is, find the rate (or rates) of output where the slope of the profit function is zero.[5]

---

[5]Sometimes the result of this optimization process is a quadratic equation, that is,

$$ax^2 + bx + c = 0$$

The solution (i.e., the value or values of $x$ for which the equation is true) is given by the general equation

$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

For example, given the quadratic equation

$$x^2 - 60x + 500 = 0$$

the solution is

$$x_1, x_2 = \frac{-(-60) \pm \sqrt{(-60)^2 - 4(1)(500)}}{2(1)}$$

$$x_1 = 50$$
$$x_2 = 10$$

$$\frac{d\pi}{dQ} = -2Q + 16 = 0$$

$$2Q = 16$$

$$Q = 8$$

But will this output rate result in a profit maximum or a profit minimum? Remember, setting the first derivative equal to zero and solving results in an extremum, but it could be a maximum or a minimum. But Figure 2A.4 shows that if the total function is maximized, the marginal function has a negative slope. Conversely, a minimum point on a total function is associated with a positive slope of the marginal function.

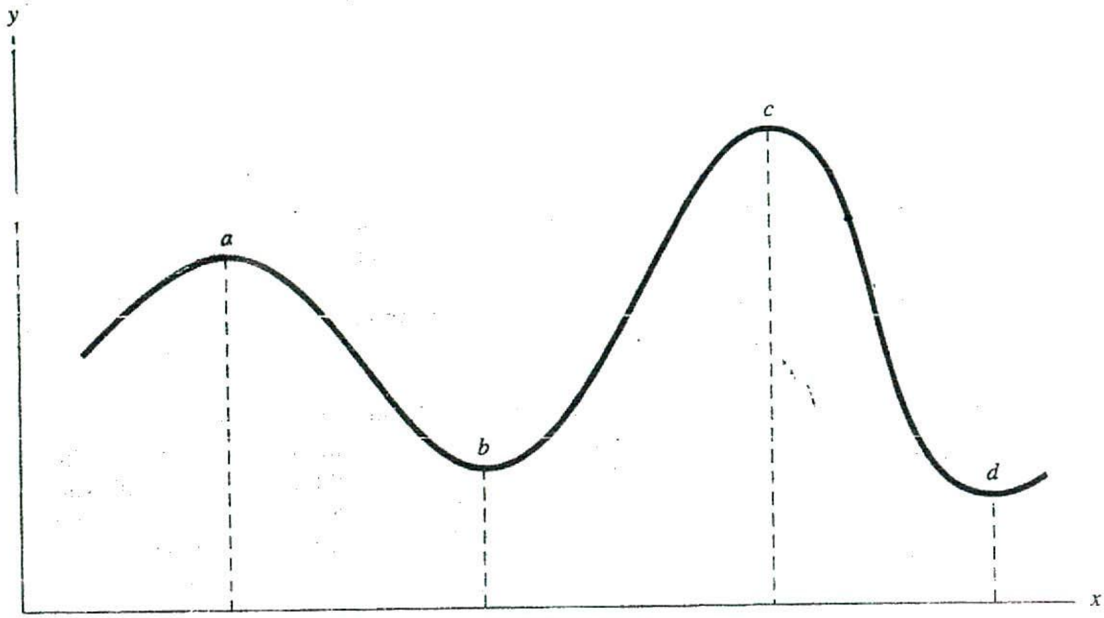
Because the slope of the marginal function is the first derivative of that marginal function, a simple test to determine if a point is a maximum or minimum is suggested. Find the second derivative of the total function and evaluate it at the point where the slope of the total function is zero. If the second derivative is negative (i.e., the marginal function is decreasing), the total function is at a maximum. If the second derivative is positive at that point (i.e., the marginal function is increasing), the total function is at a minimum point.

In the profit-maximization problem just discussed, the second derivative of the profit function is $-2$, which is negative. Therefore, profit is maximized at $Q = 8$.

When finding the extremum of any function, setting the first derivative equal to zero is called the *first-order condition,* meaning that this condition is necessary for an extremum but not sufficient to determine if the function is at a maximum or a minimum. The test for a maximum or a minimum using the second derivative is called the *second-order condition.* The first- and second-order conditions together are said to be sufficient to test for either a maximum or a minimum point. These conditions are summarized as follows:

| | *Maximum* | *Minimum* |
|---|---|---|
| First-order condition | $\frac{dy}{dx} = 0$ | $\frac{dy}{dx} = 0$ |
| Second-order condition | $\frac{d^2y}{dx^2} < 0$ | $\frac{d^2y}{dx^2} > 0$ |

In some problems there will be two or possibly more points where the first derivative is zero. Therefore, all these points will have to be evaluated using the second-order condition to test for a minimum or a maximum. As shown in Figure 2A.5, a function could have several points, such as *a, b, c,* and *d,* where the slope is zero. Points *a* and *c* are relative maxima and *b* and *d* are relative minima. The term *relative* means that the point is an extremum only for part of the function. Point *a* is a maximum relative to other points on the function around it. It is not the maximum point for the entire function because the value of *y* at point *c* exceeds that at point *a*. To find the absolute maximum, the value of the equation must be determined for all relative maxima within the range that the function is defined and also at each of the end points.




[illegible]

## THE PARTIAL DERIVATIVE

Many economic phenomena are described by multivariate functions (i.e., equations that have two or more independent variables). Given a general multivariate function such as

$$y = f(x, z)$$

the first partial derivative of $y$ with respect to $x$, denoted as $\delta y/\delta x$ or $f_x$, indicates the slope relationship between $y$ and $x$ when $z$ is held constant. This partial derivative is found by considering $z$ to be fixed and taking the derivative of $y$ with respect to $x$ in the usual way. Similarly, the partial derivative of $y$ with respect to $z$ (i.e., $\delta y/\delta z$ or $f_z$) is found by considering $x$ to be a constant and taking the first derivative of $y$ with respect to $z$.

For example, consider the function

$$y = x^2 + 3xz + z^2$$

To find the partial derivative $\partial y/\partial x$ or $f_x$, consider $z$ as a fixed and take the derivative. Thus

$$f_x = 2x^{2-1} + (3z) + 0 = 2x + 3z$$

This partial derivative means that a small change in $x$ is associated with $y$ changing at the rate $2x + 3z$ when $z$ is held constant at a specified level. For example, if $z = 2$, the slope associated with $y$ and $x$ is $2x + 3(2)$ or $2x + 6$. If $z = 5$, the slope associated with $y$ and $x$ would be $2x + 3(5)$, or $2x + 15$.

Similarly, the partial derivative of $y$ with respect to $z$ would be

$$f_z = 0 + 3x + 2z^{2-1} = 3x + 2z$$

This means that a small change in $z$ is associated with $y$ changing at the rate $3x + 2z$ when $x$ is held constant.

## OPTIMIZATION AND MULTIVARIATE FUNCTIONS

The approach to finding the maximum or minimum value of a multivariate function involves three steps. First, find the partial derivative of the function with respect to each independent variable. Second, set all the partial derivatives equal to zero. Finally, solve the system of equations determined in the second step for the values of each of the independent variables. That is, if

$$y = 4 - x^2 - 2z + xz + 2z^2$$

then the partial derivatives are

$$\frac{\partial y}{\partial x} = -2x + z$$

$$\frac{\partial y}{\partial z} = -2 + x + 4z$$

Setting these derivatives equal to zero

$$-2x + z = 0$$

$$-2 + x + 4z = 0$$

and solving these two equations simultaneously for $x$ and $z$ yields

$$x = \frac{2}{9}$$

$$z = \frac{4}{9}$$

These values of $x$ and $z$ minimize the value of the function. The approach to testing whether the optimizing solution results in a maximum or minimum for a multivariate function is complex and beyond the scope of this book. In this text, the context of the problem will indicate whether a maximum or minimum has been determined.

**Key Concepts**

- Higher-order derivatives are found by repeatedly taking the first derivative of each resultant derivative.
- The maximum or minimum point of a function [$y = f(x)$] can be found by setting the first derivative of the function equal to zero and solving for the value or values of $x$.
- When the first derivative of a function is zero, the function is at a maximum if the second derivative is negative, or at a minimum if the second derivative is positive.
- For a function having two or more independent variables [e.g., $y = f(x, z)$], the partial derivative $\partial y/\partial x$ is the slope relationship between $y$ and $x$, assuming $z$ to be held constant.
- Optimizing a multivariate function requires setting each partial derivative equal to zero and then solving the resulting system of equations simultaneously for the values of each independent variable.

## Problems

2A-1. Determine the first and second derivatives of each of the following functions.

a. $Y = 10$

b. $Y = 3X^2 + 4X + 25$

c. $Y = (X^2 - 4)(X^2 + 2X + 5)$

d. $Y = \dfrac{X^2 + 3X + 4}{X^2 - 4}$

e. $Q = 100 - 0.2P^2$

f. $R = 500Q(1 - 5Q)$

g. $Y = aX^2 + bX + c$

h. $C = 2{,}000 - 200X^2 + 3X^3$

i. $Y = (3X - 2)^2$

j. $Y = 3X^2(2X^3 - 2)^3$

2A-2. Determine all the first-order partial derivatives for each of the following functions.

a. $y = 3X^2 + 2XZ + Z^2$

b. $Q = 10K^{0.5}L^{0.6}$

c. $Q = 100P_1^{-1.2}P_2^{1.5}Y^{1.0}$

d. $C = 200 + 10X_1^2 + 2X_1X_2 + 3X_2^2$

2A-3. Given the multivariate function

$$Y = 50 + 18X + 10Z - 5XZ - 2X^2$$

determine the values of $X$ and $Z$ that maximize the function.

2A-4. The total revenue ($TR$) function for a firm is given by

$$TR = 1{,}000Q - 10Q^2$$

where $Q$ is the rate of output per period. Determine the rate of output that results in maximum total revenue. (Be sure that you have maximized, not minimized, total revenue.)

2A-5. Smith and Wesson have written a new managerial economics book for which they receive royalty payments of 15 percent of total revenue from sales of the book. Because their royalty income is tied to revenue, not profit, they want the publisher to set the price so that total revenue is maximized. However, the publisher's objective is maximum profit. If the total revenue function is

$$TR = 100{,}000Q - 10Q^2$$

and the total cost function is

$$TC = 10{,}000 + 20Q + Q^2$$

determine

a. The output rate that will maximize total royalty revenue and the amount of royalty income that Smith and Wesson would receive.

b. The output rate that would maximize profit to the publisher. Based on this rate of output, what is the amount of royalty income that Smith and Wesson would receive? Compare the royalty income of Smith and Wesson to that determined in part (a). (*Hint:* First determine a function for total profit by subtracting the cost function from the total revenue function.)

2A-6. A study indicated that the average cost function for a high school is

$$AC = 10.3 - 0.4Q + 0.00012Q^2$$

Where $Q$ is the number of students in the school

a. What size school (i.e., in terms of number of students) results in minimum average cost?

b. Find equations for total and marginal cost.

2A-7. A firm has determined that its annual profits depend on the number of salespeople it employs and the amount spent on advertising. Specifically, the relationship between profits, $\pi$ (in millions), salespeople, $S$ (in thousands), and advertising expenditures, $A$ (in millions), is

$$\pi = -10 + 60S + 10A - 2S^2 - A^2$$

Determine the number of salespeople and the amount of advertising expenditures that would maximize the firm's profits.

2A-8. A firm produces two products, milk and cheese. $Q_1$ and $Q_2$ represent the output rates for milk and cheese, respectively. The profit function is

$$\text{II} = -100 + 20Q_1 + 60Q_2 - 10Q_1^2 - 2Q_2^2 - 2Q_1Q_2$$

Determine the output rate for each product that will maximize profit.

# Integrating CASE Study I

## Olsen's Pre-Owned CD Players

Sally Olsen, a recent business school graduate, is thinking about starting her own business, and she has determined that she can procure and market recycled compact disk changers at $200 apiece. An expert in the field has offered the opinion that unit sales in any year will be 800, 1,000, or 1,200 with probabilities 0.2, 0.4, and 0.4, respectively. Given the nature of this industry, investors expect to earn a rate of return of about 12 percent per year. As the market for this product has matured, no growth is expected. This expert also thinks that realistically, Sally probably could maintain this business for about 15 years. After that, she probably would have to develop a new business.

The direct cost of purchasing each unit and refurbishing it is $100. In addition, Sally has the following annual fixed costs:

| | |
|---|---|
| Rent | $20,000 |
| Utilities | 15,000 |
| Property tax | 5,000 |
| Insurance | 5,000 |

Sally would pay herself a salary of $500 per week, which is lower than the $40,000 per year offer that she is regularly offered by several large firms, including National Electronic Sales, Inc. This salary offer is the same as that being offered to others with Sally's training and experience.

Requirements:

1. Develop accounting and economic income statements for each possible level of unit sales.
2. Based on expected profit, what is the value of this firm?
3. After Sally completed her business plan but before she actually started the operation, National Electronics increased its salary offer to Sally to $60,000 per year. Again, this is now the salary being offered to others with skills similar to Sally's. Under this condition, repeat requirements 1 and 2 as indicated above. ■

# PART II

# Demand

CHAPTER 3

*Demand Theory and Analysis*

CHAPTER 4

*Regression Techniques and Demand Estimation*

CHAPTER 5

*Business and Economic Forecasting*

INTEGRATING CASE STUDY

*Study II: Southern Turkey*

# CHAPTER 3

# Demand Theory and Analysis

## PREVIEW

Demand theory and analysis can be a source of many useful insights for business decision making. Indeed, it is difficult to overstate the importance of understanding demand. Ultimately the success or failure of a business depends primarily on its ability to generate revenues by satisfying the demands of consumers. Firms that are unable to attract the dollar votes of consumers are soon forced out of the market.

The fundamental objective of demand theory is to identify and analyze the basic determinants of consumer needs and wants. An understanding of the forces behind demand is a powerful tool for managers. Such knowledge provides the background needed to make pricing decisions, forecast sales, and formulate marketing strategies.

This chapter begins with a discussion of individual demand, market demand, and the demand faced by the firm. The focus then shifts to a basic concern of managers—the total revenue earned by the firm. Finally, the concept of elasticity is introduced as a tool for measuring the responsiveness of quantity demanded to changes in prices and income. Three elasticity measures are discussed: price elasticity, income elasticity, and cross elasticity.

## INDIVIDUAL DEMAND

Consumer choice can be a difficult task in a modern economic system. In the United States and in other industrialized nations, tens of thousands of goods and services are offered for sale. But the purchases of most consumers are constrained by their income. Many would like to drive a Ferrari, dine at the best restaurants, and vacation in Europe. However, the relevant options often are a used Ford, a pizza, and a few days at the beach.

In determining what to purchase, individual consumers face a constrained optimization problem. That is, given their income (the constraint), they select that combination of goods and services that maximizes their personal satisfaction. Implicitly, these choices involve a comparison of the satisfaction associated with having a good or service and its opportunity cost, that is, what must be given up in order to obtain it. In a market economy, opportunity costs are reflected by prices. Thus prices act as signals to guide consumer decisions. A high price denotes a significant opportunity cost, while a lower price indicates that less must be given up.

One of the most basic concepts in demand theory is the *law of demand.* In its most simple form, this law states that there is an inverse relationship between price and quantity demanded—as price increases, quantity demanded will decrease. The law of demand can be explained in terms of substitution and income effects resulting from price changes.[1] The substitution effect reflects changing opportunity costs. When the price of a good increases, its opportunity cost in terms of other goods also increases. Consequently, consumers may substitute other goods for the good that has become more expensive. The purchase of relatively more chicken and pork when beef prices increase is an example of a substitution effect.

Next consider the income effect. When the price of a good increases, the consumer's purchasing power is reduced. That is, at higher prices the individual cannot buy the same bundle of goods as before. For example, $30 will buy six pounds of chocolates at $5 per

[1]See appendix, pp. 101–111.

pound, but only five pounds at $6 per pound. The change in purchasing power is called an *income effect* because the price increase is equivalent to a reduction in the consumer's income.

## Case Study

### The Law of Demand in Perkasie, Pennsylvania

Among the many environmental problems facing the United States is how to dispose of the vast amounts of garbage generated each day by households and businesses. In 1960, Americans discarded an average of 2.6 pounds of trash per person per day, but today the number is 3.8 pounds. As the volume of garbage grows, existing disposal sites are filling up, and it is becoming increasingly difficult to find new locations near urban areas for landfills.

One small community used the law of demand to ease its garbage collection problem. Residents of Perkasie, Pennsylvania, were paying an annual fixed fee of $120 per resident for garbage collection and discarding a daily average of 2.2 pounds of trash per person. Because the collection fee was fixed, the additional expense to residents of trash disposal was zero, and they had no financial incentive to conserve on the amount of trash they produced.

Perkasie began charging by the bag for garbage collection. The city required that all trash be placed in special bags sold by the city. For example, a large bag had a capacity of about 40 pounds and sold for $1.50. Thus, the marginal cost to residents of generating additional trash increased from zero to about four cents per pound. Garbage that was not in an approved bag was not picked up. In addition, the city introduced a recycling program. Each household was given buckets to be filled with cans and bottles that were picked up every week. The city also arranged for newspapers to be collected once a month.

The result was predictable—people began to dump less trash. During the first year the program was in effect, trash collections per person declined to less than one pound per day. Perkasie citizens benefited because they paid 30 percent less than before, and the city reduced its garbage collection costs by 40 percent. ■

## MARKET DEMAND

Although choices by individuals are the basis of the theory of demand, it is total or market demand that is of primary interest to managers. The market demand for a good or service is the sum of all individual demands. For example, consider a market that consists of only two buyers. The demand curves for these two consumers are depicted in Figure 3.1. These demand curves show the relationship between price and quantity demanded. Consumer 1's demand curve is shown in the first panel ($D_1D_1$) and that of consumer 2

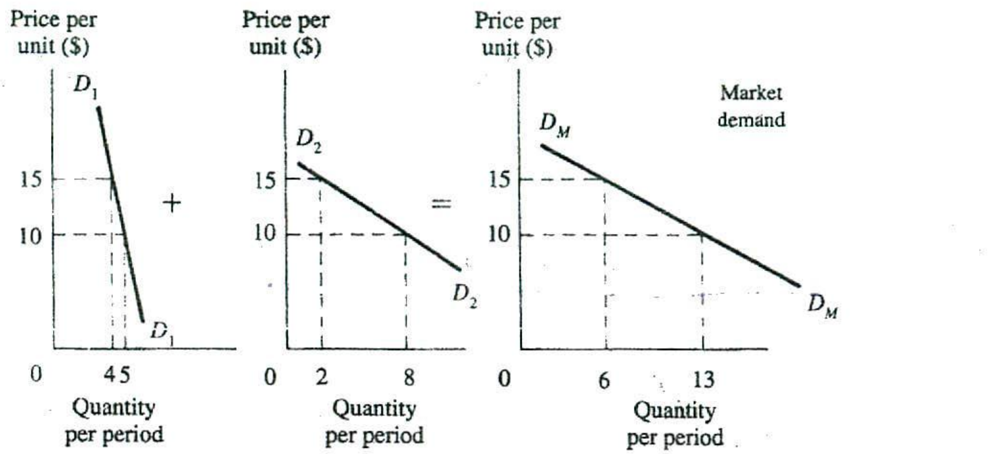



FIGURE 3.1 The Market Demand Curve

in the second panel ($D_2D_2$). At a price of \$10, the individual quantities demanded are 5 and 8 units, respectively. Hence the total market demand ($D_MD_M$, as shown in the third panel) is 13 units. The market demand at any price is the sum of the individual quantities demanded at that price.

Graphically, the market demand curve is the horizontal summation of the individual demand curves. That is, for any given price, the market demand curve is the sum of the horizontal distances from the vertical axis to each individual demand curve.

## Determinants of Market Demand

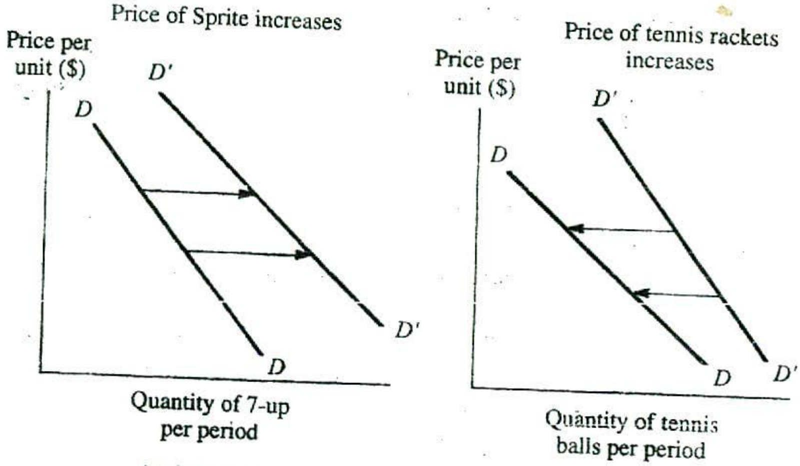
The effect of a change in price is depicted as a movement from one point to another *along* a particular demand curve. For example, in Figure 3.1, as price increases from \$10 to \$15, moving along the market demand curve $D_MD_M$ depicts a decrease in quantity demanded from 13 to 6 units. A movement along the demand curve in response to a change in the price of a good or service is referred to as a *change in quantity demanded.*

The price of a good or service is not the only determinant of demand. However, in plotting a demand curve, it is assumed that other factors that affect demand are held constant. When these factors are allowed to vary, the demand curve will shift. Such shifts are referred to as *changes in demand.* A shift to the right is called an increase in demand, meaning that consumers demand more of the good or service at each price than they did before. A leftward shift indicates a decrease in demand. That is, less is demanded at each price than before.

In a market economy, firms must be responsive to consumer demands. Thus, it is important that managers understand the determinants of demand. Some of the most important are consumer preferences, income levels, and prices of other goods.[2]

**Consumer Preferences** Obviously, an important determinant of demand is the preferences of consumers. These preferences can change rapidly in response to advertising.

[2]Other factors, such as population, expectations, and government policies, can also affect demand. Although they may be important, for ease of exposition they are not considered here.




(a) Substitutes (b) Complements

Demand Curves for Substitutes and Complements

fads, and customs. There was a time when gloves were considered a must for the well-dressed woman. Today, gloves are usually worn only on special occasions. This change in preferences caused a decrease in the demand for gloves, meaning that fewer gloves are demanded at any given price. In contrast, pants for women have become more popular in recent years. In this case, the shift in preferences resulted in an increase in demand, with more women's pants now demanded at each price.

**Income** Demand is also affected by the amount of income that consumers have available to spend. For most goods, an increase in consumer income would cause the demand curve for the product to shift to the right. For example, a few years ago, revenues from oil and gas leases left the state of Alaska with surplus revenues. The state responded by providing a cash grant to every resident of the state. Alaskans used much of this extra income to buy additional goods and services. That is, the increased income resulted in an increase in demand for many goods and services. Later, lower oil prices reduced incomes in Alaska. The result was a decrease in demand for many goods and services.

**Prices of Other Goods** The demand for a good is often influenced by changes in the prices of other goods. The nature of the impact depends on whether the goods are substitutes or complements. *Substitutes* are goods that have essentially the same use. When the price of a good goes up, the demand for its substitutes is likely to increase. For example, 7-Up and Sprite are similar lemon-flavored soft drinks. An increase in the price of Sprite would cause people to purchase less of that beverage and consume more 7-Up. Thus the demand curve for 7-Up would shift to the right from *DD* to *D'D'*, as shown in Figure 3.2a. Note that more 7-Up is demanded at each price than before the shift.

Goods that are often used together are called *complements*. An increase in the price of one such good will cause the demand for its complement to decrease. Consider tennis rackets and balls. If the price of rackets increases, fewer people will play tennis. With fewer people involved in the sport, fewer tennis balls will be purchased. This outcome is illustrated by the leftward shift of the demand curve from *DD* to *D'D'* in Figure 3.2b.

## The Market Demand Equation

The market demand function can also be expressed mathematically. If the primary determinants of demand are the price of the product, income, consumer preferences, and the prices of other goods and services, the demand equation can be written as

$$Q_D = f(P, I, P_o, T) \tag{3-1}$$

where $P$ is the price of the good or service, $I$ is income, $P_o$ represents the prices of other goods, and $T$ is a measure of consumer tastes and preferences. Equation (3-1) suggests that there is a correspondence between the quantity demanded and the variables on the right-hand side. However, the equation implies only that there are general relationships. It says nothing about their nature and magnitude. For example, equation (3-1) provides no information about how quantity demanded would be affected by an increase in income. Quantifying this information requires that a functional form be chosen to represent the equation for market demand. The linear form is shown below.

$$Q_D = B + a_pP + a_I I + a_o P_o + a_T T \tag{3-2}$$

The coefficients $a_p$, $a_I$, $a_o$, and $a_T$ indicate the change in quantity demanded of one-unit changes in the associated variables. For example, $a_p$ is the coefficient of price. Its interpretation is that, holding the other three variables constant, quantity demanded changes by $a_p$ units for each one-unit change in price. In most cases, $a_p$ will be negative. To illustrate, if $a_p = -2$ and price is measured in dollars, a $1 increase in price would be associated with a two-unit decrease in quantity demanded.

For many purposes it is useful to focus on the relationship between quantity demanded and the price of the good or service while holding the other variables constant. If $I$, $P_o$, and $T$ are not allowed to vary, then demand is a function only of $P$. Hence the linear form of the demand equation can be written as

$$Q_D = B + a_pP \tag{3-3}$$

where $B$ represents the combined influence of all the other determinants of demand and $a_p \leq 0$. This simple demand equation is the basis for much of the analysis in the remainder of the chapter.

---

**Example** 

### The Market Demand for Tests

Max, a graduating senior, has accumulated an impressive file of tests during his college career. But now he needs to sell his test collection to obtain money for his impending marriage. Three wealthy friends express interest in buying some of the tests. Max determines that their individual demand equations are as follows:

$$Q_1 = 30.00 - 1.00P$$
$$Q_2 = 22.50 - 0.75P$$
$$Q_3 = 37.50 - 1.25P$$

where the quantity subscripts denote each of the three friends and price is measured in dollars per test.

What is the market demand equation for Max's tests, and how many more tests can he sell for each one-dollar decrease in price? If he has a file of 60 tests, what price should he charge to sell his entire collection?

**Solution** Market demand, $Q_m$, is the sum of the individual demands. Thus

$$Q_m = Q_1 + Q_2 + Q_3 = (30.00 - 1.00P) + (22.50 - 0.75P) + (37.50 - 1.25P)$$

Simplifying yields

$$Q_m = 90.00 - 3.00P$$

Because $P$ is measured in dollars, a one-dollar decrease in price will increase quantity demanded by three tests. To sell the entire 60-test collection, the price must be set such that $Q_m = 60$. That is,

$$60 = 90 - 3.00P$$

Solving this equation gives $P = \$10.00$. Substituting this price back into the individual demand equations gives $Q_1 = 20$, $Q_2 = 15$, and $Q_3 = 25$.

## Market Demand versus Firm Demand

Thus far, the discussion has focused on market demand curves. But from the perspective of managers, it is the demand curve facing the individual firm that is most relevant in pricing and output decisions.

Where a firm is the only seller in a market, the relevant demand curve is the market demand curve. Consequently, the firm will bear the entire impact of changes in incomes, consumer preferences, and prices of other goods. Similarly, the pricing policies of the firm will have a significant impact on purchasers of the firm's product. But few businesses are the only sellers of a good or service. In the vast majority of cases, a firm supplies only part of the total market. Consequently, the demand curve faced by the individual firm is not the same as the market demand curve.

One major difference between firm and market demand is that additional factors affect demand at the firm level. Perhaps the most important factor is decisions made by competitors. For example, a price cut by one firm probably would decrease sales of rival firms unless those firms also reduced price. Similarly, an effective advertising campaign can increase a firm's sales at the expense of its competitors.

Another difference between firm and market demand is the quantitative impact of changes in tastes, income, and prices of other goods. Consider beef and pork as examples of substitute goods. Suppose the market demand equation indicates that a 1-cent per pound increase in the price of beef would increase the market demand for pork by 1 million pounds per year. Now consider the demand equation for a small meatpacker with a 1 percent market share of pork sales. Because the firm's share is a small fraction of the total market, the impact on the firm's sales resulting from the change in beef prices will be much less than the total market change. That is, if the demand equation for the small firm was estimated, the coefficient showing the effect of changes in the price of beef would be much smaller than the coefficient for the market demand equation. Similarly, the coefficients showing the effects of changes in tastes and incomes would also be less than for the market demand equation.

For the remainder of this chapter, the discussion will focus on the demand curve faced by the individual firm. However, most of the concepts discussed are equally applicable to market demand curves.

**Key Concepts**

- A change in quantity demanded refers to a movement along a demand curve caused by a change in the price of the good or service.
- A change in demand is represented by a shift of the demand curve resulting from a change in consumer preferences, income, or prices of other goods.
- If two goods are substitutes, when the price of one good increases, demand for the other also increases.
- If two goods are complements, a higher price for one will cause a decrease in the demand for the other.

## TOTAL AND MARGINAL REVENUE

One indication of a firm's success is the total revenue generated by the sale of its products. Rankings of firm size are usually made on the basis of total revenue. Similarly, growth is often expressed in terms of increases in total revenue. In that it reflects the ability of the firm to satisfy consumer demands, the use of total revenue as a measure of success has some merit.

The first two columns of Table 3.1 provide information about the demand faced by a firm. By multiplying price times quantity, the total revenue associated with each price—quantity pair is determined. These data are also shown in Table 3.1. Note that total revenue increases as price goes from $1 to $5 and then decreases for prices greater than $6. This suggests that sound pricing decisions require information about demand. In some cases higher prices may increase total revenue, whereas in other circumstances a price increase can have the opposite effect.

Table 3.1 also shows marginal revenue. Marginal revenue is defined as the change in total revenue associated with the sale of one more unit of the product. For example, as quantity goes from four to five, total revenue increases from $28 to $30. Hence, the

[illegible]

| *Price* | *Quantity* | *Total Revenue* | *Marginal Revenue* |
|---|---|---|---|
| $10 | 1 | $10 | — |
| 9 | 2 | 18 | $8 |
| 8 | 3 | 24 | 6 |
| 7 | 4 | 28 | 4 |
| 6 | 5 | 30 | 2 |
| 5 | 6 | 30 | 0 |
| 4 | 7 | 28 | −2 |
| 3 | 8 | 24 | −4 |
| 2 | 9 | 18 | −6 |
| 1 | 10 | 10 | −8 |

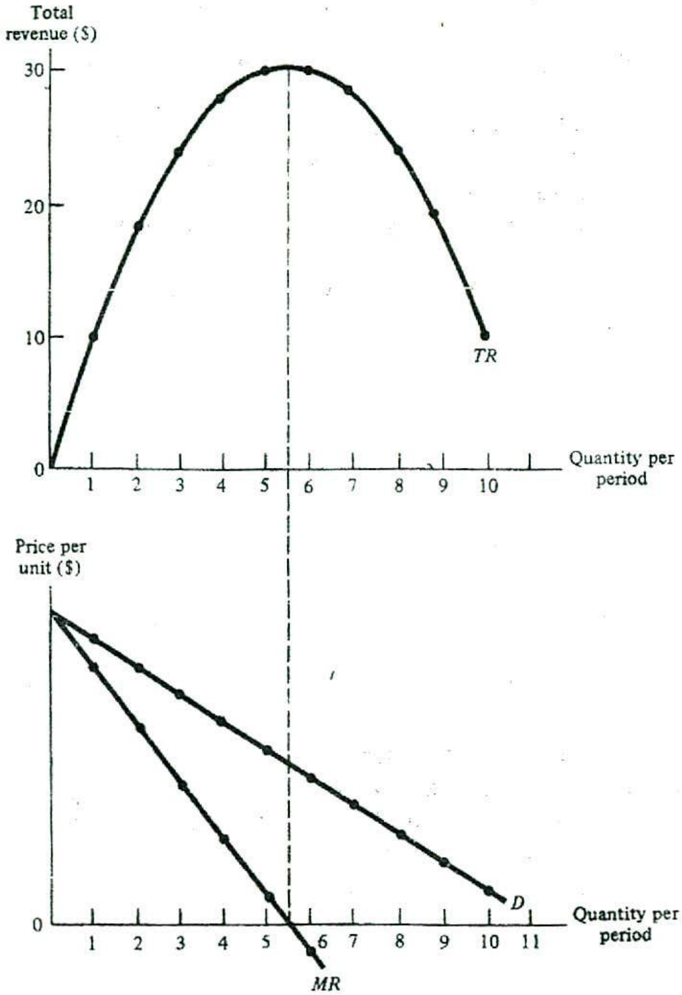



FIGURE 3.3 Total Revenue, Marginal Revenue, and the Demand Curve

marginal or extra revenue associated with the fifth unit is $2. Note that marginal revenue declines as quantity increases. Beyond six units, marginal revenue is negative. The explanation stems from the inverse relationship between price and quantity. To sell extra units, the firm must reduce the price of all the units sold. Negative marginal revenue means that the dollars received from selling the extra unit are not sufficient to compensate for the dollars lost as a result of selling all other units at a lower price. Clearly, a firm should not increase output beyond the point where marginal revenue is zero.

The total and marginal revenue data of Table 3.1 can be plotted on a graph. If fractional units are allowed, the line has the appearance of a smooth curve, as shown in Figure 3.3. Note the relationship between the total and marginal revenue curves. As long as total revenue is increasing, marginal revenue is positive. At the maximum point on the total revenue curve, marginal revenue is zero. But beyond that point, marginal revenue is negative.

Figure 3.3 also shows the relationship between the marginal revenue curve and the demand curve. Note that the two curves intercept the price axis at the same point. Using calculus, it can easily be shown that for linear demand equations, the value of the slope of the marginal revenue curve is twice the value of the slope of the associated demand curve. Because the two curves have the same price intercept, this implies that the quantity intercept of the marginal revenue curve is exactly half that of the demand curve.

The marginal revenue equation can be derived from the demand equation. Suppose that the demand equation is given by $Q = B + a_pP$, where $a_p \leq 0$. Solving for price, the demand equation becomes

$$P = \frac{-B}{a_p} + \frac{Q}{a_p} \tag{3-4}$$

Multiplying by quantity gives

$$TR = PQ = -\frac{B}{a_p}Q + \frac{Q^2}{a_p} \tag{3-5}$$

Marginal revenue is the derivative of $TR$ with respect to $Q$. Thus

$$MR = \frac{d(PQ)}{dQ} = -\frac{B}{a_p} + \frac{2Q}{a_p} \tag{3-6}$$

Note that the marginal revenue equation has the same intercept $(-B/a_p)$ as the demand equation and that the slope of marginal revenue, $2/a_p$, is twice the slope of the demand equation, $1/a_p$. Also note that $MR = 0$ at $Q = B/2$. This point corresponds to the maximum value of the total revenue function. For $Q < B/2$, marginal revenue is positive and total revenue is increasing. For $Q > B/2$, marginal revenue is negative and total revenue is decreasing.

**Key Concepts**

- Marginal revenue is the change in revenue associated with a one-unit change in output.
- Marginal revenue is zero at the quantity that generates maximum total revenue and is negative beyond that point.
- For a linear demand curve, the absolute value of the slope of the marginal revenue curve is twice that of the demand curve.

## PRICE ELASTICITY

In the preceding section it was demonstrated that higher prices do not always result in greater total revenue. A price change can either increase or decrease total revenue, depending on the nature of the demand function. The uncertainty involved in pricing decisions could be reduced if managers had a method of measuring the probable effect of price changes on total revenue. One such measure is *price elasticity of demand,* which is defined as the percentage change in quantity demanded, divided by the percentage change in price. That is,

$$E_p = \frac{\%\Delta Q}{\%\Delta P} \tag{3-7}$$

where the symbol $\Delta$ is used to denote change. Thus, price elasticity indicates the percentage change in quantity demanded for a 1 percent change in price.

For example, suppose that a firm increases the price of its product by 2 percent and quantity demanded subsequently decreases by 3 percent. The price elasticity would be

$$E_p = \frac{-3\%}{2\%} = -1.5$$

Notice that $E_p$ is negative. Except in rare and unimportant cases, price elasticity is always less than or equal to zero.[3] The explanation is the law of demand, which states that price and quantity demanded are inversely related. Thus, when the price change is positive, the change in quantity demanded is negative, and vice versa.

## Point versus Arc Elasticity

There are two approaches to computing price elasticities. The choice between the two depends on the available data and the intended use. Arc elasticities are appropriate for analyzing the effect of discrete (i.e., measurable) changes in price. For example, a price increase from $1 to $2 could be evaluated by computing the arc elasticity. In actual practice, most elasticity computations involve the arc method. The other choice is point elasticity. This approach can be used to evaluate the effect of very small price changes or to compute the price elasticity at a particular price. Point elasticities are important in theoretical economics.

**Arc Elasticity** The percentage change in price is the change in price divided by price (i.e., $\Delta P/P$). Similarly, the percentage change in quantity demanded is the change in quantity divided by quantity (i.e., $\Delta Q/Q$). Thus price elasticity can be expressed as

$$E_p = \frac{\Delta Q/Q}{\Delta P/P} \tag{3-8}$$

By rearranging terms, equation (3-8) can be written as

$$E_p = \frac{\Delta Q}{\Delta P}\frac{P}{Q} \tag{3-9}$$

Note that $P$ and $Q$ specify a particular point on the demand curve, while $\Delta Q/\Delta P$ is the reciprocal of the slope of the demand curve. Thus the price elasticity will vary depending on where the measurement is taken on the demand curve and also with the slope of the demand curve.

[3] In some texts, price elasticities are multiplied by $-1$ to make them positive. This convention is not adopted here because it obscures the inverse relationship between price and quantity demanded.

The price and quantity information presented earlier in the chapter can be used to demonstrate the calculation of the arc price elasticity. Table 3.1 on page 78 indicates that at a price of \$6, quantity demanded is 5 units. The table also shows that if the price increases to \$10, quantity demanded will be 1 unit. Thus for a \$4 increase in price, the change in quantity demanded is −4 units. That is,

$$\frac{\Delta Q}{\Delta P} = \frac{-4}{4} = -1$$

Determining the value of $P/Q$ poses a problem. Which should be used, the price−quantity data before the price change or the price–quantity values after the price increase? The choice makes a significant difference in the computed price elasticity. For example, if the initial data are used, then

$$E_p = \frac{\Delta Q}{\Delta P}\frac{P}{Q} = -1 \times \frac{6}{5} = -1.20$$

However, if the price−quantity data after the price change are selected, then

$$E_p = \frac{\Delta Q}{\Delta P}\frac{P}{Q} = -1 \times \frac{10}{1} = -10.00$$

In the first case, the price elasticity estimate indicates that a 1 percent increase in price results in a quantity decline of just over 1 percent. But the second estimate implies that the percentage impact on quantity demanded is 10 times that of the percentage price change.

The conventional approach used to calculate arc elasticities is to use average values for price and quantity. Thus arc price elasticity is defined as

$$E_p = \frac{Q_2 - Q_1}{(Q_2 + Q_1)/2} \Big/ \frac{P_2 - P_1}{(P_2 + P_1)/2} \qquad (3\text{-}10)$$

where $P_1$ and $Q_1$ are the initial price−quantity pair and $P_2$ and $Q_2$ are the price−quantity values after the price change. Simplifying and rearranging terms, equation (3-10) can be written as

$$E_p = \frac{Q_2 - Q_1}{P_2 - P_1} \times \frac{P_2 + P_1}{Q_2 + Q_1} \qquad (3\text{-}11)$$

Now using the data from Table 3.1, the arc elasticity is computed to be

$$E_p = \frac{1-5}{10-6} \times \frac{10+6}{1+5} = -2.67$$

Thus as price increases from \$6 to \$10, the average change in quantity demanded per percent change in price is −2.67 percent.

**Example** **The Price Elasticity of Demand for Playing Cards**

Suppose the market demand for playing cards is given by the equation

$$Q = 6{,}000{,}000 - 1{,}000{,}000P$$

where $Q$ is the number of decks of cards demanded each year and $P$ is the price in dollars. For a price increase from \$2 to \$3 per deck, what is the arc price elasticity?

**Solution** Using the market demand equation, the quantity demanded is 4,000,000 decks of cards at a price of \$2. Similarly, it is determined that the quantity demanded is 3,000,000 decks at a price of \$3 per deck. Note that in equation (3-11), $(Q_2 - Q_1)/(P_2 - P_1)$ is just $\Delta Q/\Delta P$. The market demand equation can be used to determine $\Delta Q/\Delta P$. In that each \$1 increase in price causes a 1,000,000 decrease in quantity demanded, it is known that $\Delta Q/\Delta P = -1{,}000{,}000$. Thus the arc price elasticity is

$$E_p = -1{,}000{,}000 \times \frac{2 + 3}{4{,}000{,}000 + 3{,}000{,}000} = -0.71$$

This means that a 1 percent increase in price will reduce the quantity demanded by 0.71 percent.

---

**Point Elasticity** Now let's consider extremely small changes in price. For $\Delta P$ approaching zero, the term $\Delta Q/\Delta P$ can be written as $dQ/dP$, where $dQ/dP$ is the derivative of $Q$ with respect to $P$. Basically, $dQ/dP$ expresses the rate at which $Q$ is changing for very small changes in $P$. For a linear demand equation, $dQ/dP$ is constant. For example, in the demand equation $Q = B + a_pP$, the derivative $dQ/dP$ is $a_p$. Thus the rate of change for a small price change is the same as for large changes. Hence for linear demand equation, $dQ/dP = \Delta Q/\Delta P$.

For small price changes, $P_1$ and $P_2$ are approximately equal. Thus either price can be used in the calculation with no significant effect on the computed price elasticity. Hence the equation for point elasticity can be written as

$$E_p - \frac{dQ}{dP}\frac{P}{Q} \tag{3-12}$$

Equation (3-12) is used to calculate price elasticities at a particular point on the demand curve. For example, consider the price–quantity data used earlier in this section. It has already been determined that $dQ/dP = -1$. If price is \$6 and the quantity demanded is 5 units, the price elasticity is

$$E_p = \frac{dQ}{dP}\frac{P}{Q} = -1 \times \frac{6}{5} = -1.20$$

That is, for very small price changes above or below \$6, the percentage change in quantity demanded is $-1.20$ times the percentage change in price. Again, the minus sign denotes that there is an inverse relationship between price and quantity.

Point elasticities can also be computed from a demand equation. Suppose that the demand equation is as follows:

$$Q_D = 100 - 4P$$

Because the relationship is linear, $dQ/dP$ is constant and equal to the rate of change in $Q_D$ for each one-unit change in $P$. Note that quantity demanded changes by $-4$ units for each unit increase in $P$. Thus $dQ/dP = -4$.

Suppose that $P = \$10$. Substituting this value into the demand equation yields $Q_D = 60$. Thus the point elasticity at $P = \$10$ is

$$E_p = -4 \times \frac{10}{60} = -0.67$$

The interpretation is that 1 percent increase in price causes a 0.67 percent reduction in the quantity demanded.

Consider a second example using the same demand equation. Suppose that $P =$ \$20, which implies that $Q = 20$. The equation is linear, so $dQ/dP = -4$, as before. Thus

$$E_p = -4 \times \frac{20}{20} = -4.0$$

That is, at $P =$ \$20, a small change in price generates a percentage change in quantit demanded that is four times the percentage change in price.

## Price Elasticity and Marginal Revenue

Table 3.2 reproduces the price, quantity, total revenue, and marginal revenue informa tion from Table 3.1. It also shows the point price elasticity at each price. Note that th absolute value of the elasticities becomes smaller as prices decrease. At $P =$ \$10, th elasticity is $-10.00$, while at $P =$ \$1, it is $-0.10$.

Often, it is useful to classify demand relationships on the basis of price elasticiti. The following classification scheme is frequently used:

| *If:* | *Then Demand Is Said to Be:* |
|---|---|
| $E_p < -1$ | Elastic |
| $E_p = -1$ | Unitary elastic |
| $-1 < E_p \leq 0$ | Inelastic |

Thus, based on the data in Table 3.2, at some price between \$5 and \$6, demand is u tary elastic because the price elasticity is equal to $-1$. For prices below this price, d mand is inelastic. For higher prices, demand is elastic.

Figure 3.4 shows the demand and marginal revenue curves for the price–quanti data of Table 3.2. The figure has been labeled to show where demand is elastic, unita elastic, and inelastic. The figure indicates that demand becomes less elastic at low prices. This is a characteristic of linear demand curves. Because the curve is linea $dQ/dP$ is a constant. Thus price elasticity is determined by the value of $P/Q$. But as pric decreases, $P/Q$ also decreases. Consequently, the absolute value of $E_p$ becomes smalle and demand becomes less elastic.

| *Price* | *Quantity* | *Price Elasticity* | *Total Revenue* | *Marginal Revenue* |
|---|---|---|---|---|
| \$10 | 1 | −10.00 | \$10 | — |
| 9 | 2 | −4.50 | 18 | \$8 |
| 8 | 3 | −2.67 | 24 | 6 |
| 7 | 4 | −1.75 | 28 | 4 |
| 6 | 5 | −1.20 | 30 | 2 |
| 5 | 6 | −0.83 | 30 | 0 |
| 4 | 7 | −0.57 | 28 | −2 |
| 3 | 8 | −0.38 | 24 | −4 |
| 2 | 9 | −0.22 | 18 | −6 |
| 1 | 10 | −0.10 | 10 | −8 |

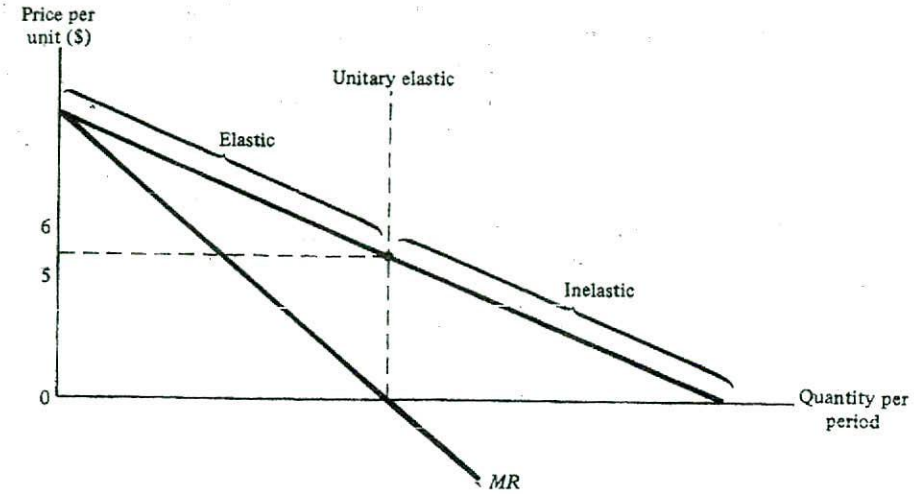



FIGURE 3.4 Price [illegible]

Figure 3.4 can be used to interpret the three elasticity categories. The figure shows that the point of unitary elasticity corresponds to the point where the marginal revenue crosses the quantity axis. That is, marginal revenue is zero where demand is unitary elastic. The explanation is not difficult. Unitary elasticity means that a 1 percent increase in price causes quantity demanded to decrease by 1 percent. But total revenue is computed by multiplying price times quantity. Thus, if demand is unitary elastic, the increase in price is exactly offset by the decrease in quantity demanded. As a result, there is no change in total revenue and marginal revenue is zero.

Figure 3.4 also shows that marginal revenue is positive where demand is elastic, and negative when demand is inelastic. Although the demand curve in Figure 3.4 is linear, these relationships are also true for nonlinear demand curves. The point where marginal revenue is zero always divides the elastic and inelastic regions of the demand curve.[4]

---

[4]However, it should be noted that there are some demand relationships for which marginal revenue is never zero. For example, consider a demand equation of the following form:

$$Q = P^a$$

Taking the derivative with respect to $P$ yields

$$\frac{dQ}{dP} = aP^{a-1}$$

Thus the price elasticity is

$$E_p(aP^{a-1})\frac{P}{Q} = \frac{aP^a}{Q}$$

But $P^a = Q$. Thus $E_p = a$. Because $a$ is a constant, it follows that $E_p$ is constant for all values of price and quantity. If $a < -1$, demand is elastic for all points on the curve. Thus marginal revenue is always positive. Conversely, if $a > -1$, demand is inelastic and marginal revenue negative. If $a = -1$, marginal revenue is zero and demand is unitary elastic for all price–quantity combinations.

The choice of the terms *inelastic* and *elastic* is appropriate when viewed in light of the relationship between price elasticity and marginal revenue. The word *inelastic* carries the connotation of something that is not flexible or responsive. Consider a vertical demand curve. For such a curve, quantity demanded is not affected by changes in price. That is, $dQ/dP = 0$ and thus $E_p = 0$. Such curves are sometimes referred to as completely inelastic.

In contrast, for a horizontal demand curve, quantity demanded is highly responsive to changes in prices. In fact, an arbitrarily small change in price causes an infinitely large change in quantity demanded. That is, $dQ/dP = -\infty$ and hence $E_p = -\infty$. Horizontal demand curves are said to be infinitely elastic. Although there are probably no goods or services for which market demand is infinitely elastic or completely inelastic, an understanding of these cases is useful in economic analysis. Also, individual producers may face demand curves that approach these extremes.

**Key Concepts**

- Price elasticity is defined as the percentage change in quantity demanded that results from a 1 percent change in price.
- Point elasticity is used where the price change is very small. Arc elasticity is appropriate for a larger change in price.
- Demand is elastic if $E_p < -1$, inelastic if $-1 < E_p \leq 0$, and unitary elastic if $E_p = -1$.
- Demand is elastic where marginal revenue is positive, unitary elastic when marginal revenue is zero, and inelastic if marginal revenue is negative.

## Determinants of Price Elasticity

Price elasticities have been estimated for many goods and services; Table 3.3 provides some examples. The short-run elasticities reflect periods of time that are not long enough for the consumer to adjust completely to changes in prices. The long-run values refer to situations where consumers have had more time to adjust.

Note the variation in elasticities in Table 3.3. The long-run demand for foreign travel by U.S. residents is elastic (i.e., $E_p = -4.10$). In contrast, the long-run demand for water is highly inelastic (i.e., $E_p = -0.14$). Electricity demand is inelastic in the short run, but elastic in the long run. In general, three factors determine the price elasticity of demand. They are (1) availability of substitutes, (2) proportion of income spent on the good or service, and (3) length of time.

**Availability of Substitutes** Products for which there are good substitutes tend to have higher price elasticities than products for which there are few adequate substitutes. Motion pictures are a good example. Movies are a form of recreation, but there are many alternative recreational activities. When ticket prices at the movie theater increase, these substitute activities replace movies. Thus, the demand for motion pictures is relatively elastic, as shown in Table 3.3.

**TABLE 3.3 Estimates of Price Elasticity**

| *Good or Service* | *Estimated Price Elasticity* | |
|---|---|---|
| Electricity | −0.13 | Short run |
| Electricity | −1.89 | Long run |
| Water | −0.14 | Long run |
| Motion pictures | −3.69 | Long run |
| Gasoline | −0.15 | Short run |
| Gasoline | −0.78 | Long run |
| Foreign travel | −4.10 | Long run |

*Sources:* H. S. Houthakker and L. D. Taylor, *Consumer Demand in the United States; Analysis and Projections* (Cambridge, Mass.: Harvard University Press, 1970), pp. 166–167; and J. L. Sweeney, "The Demand for Gasoline: A Vintage Capital Model," Department of Engineering Economics, Stanford University, 1975.

At the other extreme, consider the short-run demand for electricity. When the local utility increases prices, consumers have few options. The owner of an electrically heated home can turn down the thermostat, but the savings are limited by the desire to keep warm. The owner of an electric clothes dryer may try drying clothes on an outside line, but this option will not be very successful during the winter or for the residents of some apartment buildings. Similarly, there are not many short-run alternatives to using electricity for refrigeration and lighting. Hence the short-run demand for electricity is relatively inelastic.

**Proportion of Income Spent** Demand tends to be inelastic for goods and services that account for only a small proportion of total expenditures. Consider the demand for salt. A one-pound container of salt will meet the needs of the typical household for months and costs only a few cents. If the price of salt were to double, this change would not have a significant impact on the family's purchasing power. As a result, price changes have little effect on the household demand for salt. In contrast, demand will tend to be more elastic for goods and services that require a substantial portion of total expenditures.

**Time Period** Demand is usually more elastic in the long run than in the short run. The explanation is that, given more time, the consumer has more opportunities to adjust to changes in prices. Table 3.3 indicates that the long-run elasticity for electricity is more than ten times the short-run value. As indicated previously, in the short run, people living in an electrically heated home have few options to reduce electricity consumption. But over a longer time period, they may switch to gas or improve the energy efficiency of the home. Similarly, higher electricity prices may ultimately cause consumers to find other energy sources for cooking and clothes drying.

# *Case Study*

## The Short-Run Versus Long-Run Demand for Gasoline

As shown in the following table, gasoline prices increased dramatically from 1973 to 1981. At first, consumers had little choice but to use about the same amount of gasoline and pay the higher prices. Some vacation trips were canceled and many commuters started going to work in buses or car pools, but the options for relief were limited. From 1973 to 1975, average fuel consumption per vehicle declined from 736 to 685 gallons per year, a decrease of 7 percent. However, given more time to adjust, consumers were able to reduce the impact of higher gas prices. Smaller, fuel-efficient cars became popular, and the average miles per gallon of gasoline for passenger cars increased from 13.3 in 1973 to 15.7 in 1981. People also changed jobs or moved closer to their places of work. These and other changes in driving habits reduced the average number of miles driven per car from 9,800 to 8,700 over the same period. The net effect of these changes was that fuel consumption per vehicle in the United States declined from 736 to 555 gallons per year between 1973 and 1981, a reduction of nearly 25 percent. Clearly, the long-run demand for gasoline was more elastic than the short-run demand. ■

**Gasoline Prices and Consumer Response**

| *Year* | *Average Price of Gasoline* | *Average Miles per Gallon* | *Average Miles Driven per Vehicle per Year* | *Average Fuel Consumption (gallons)* |
|---|---|---|---|---|
| 1973 | $0.40 | 13.3 | 9,800 | 736 |
| 1975 | 0.57 | 13.7 | 9,400 | 685 |
| 1977 | 0.62 | 14.1 | 9,600 | 680 |
| 1979 | 0.86 | 14.5 | 9,300 | 638 |
| 1981 | 1.31 | 15.7 | 8,700 | 555 |

*Source:* U.S. Department of Commerce, Bureau of the Census. *Statistical Abstract of the United States* (Washington, D.C.: U.S. Government Printing Office, selected years).

### Price Elasticity and Decision Making

Information about price elasticities can be extremely useful to managers as they contemplate pricing decisions. If demand is inelastic at the current price, a price decrease will result in a decrease in total revenue. Alternatively, reducing the price of a product with elastic demand would cause revenue to increase.[5] The effect on total revenue would be the reverse for a price increase. However, if demand is unitary elastic, price changes will not change total revenues.

[5]However, a price reduction is not always the correct strategy when demand is elastic. The decision must also take into account the impact on the firm's costs and profits. More will be said about pricing strategy in later chapters.

The relationship between elasticity and total revenue can be shown using calculus. Total revenue is price times quantity. Taking the derivative of total revenue with respect to quantity yields marginal revenue:

$$MR = \frac{d(TR)}{dQ} = \frac{d(PQ)}{dQ} = P + Q\frac{dP}{dQ} \tag{3-13}$$

Equation (3-13) states that the additional revenue resulting from the sale of one more unit of a good or service is equal to the selling price of the last unit ($P$), adjusted for the reduced revenue from all other units sold at a lower price ($QdP/dQ$). This equation can be written

$$MR = P\left(1 + \frac{Q\,dP}{P\,dQ}\right) \tag{3-14}$$

But note that $(Q/P)\,dP/dQ = 1/E_p$. Thus

$$MR = P\left(1 + \frac{1}{E_p}\right) \tag{3-15}$$

Equation (3-15) indicates that marginal revenue is a function of the elasticity of demand. For example, if demand is unitary elastic, $E_p = -1$ then

$$MR = P\left(1 + \frac{1}{-1}\right) = 0$$

Because marginal revenue is zero, a price change would have no effect on total revenue. In contrast, if demand is elastic, $E_p < -1$ and $(1 + 1/E_p) > 0$. Hence, marginal revenue is positive, which means that, by increasing quantity demanded, a price reduction would increase total revenue. Equation (3-15) also implies that if demand is inelastic, marginal revenue is negative, indicating that a price reduction would decrease total revenue.

Some analysts question the usefulness of elasticity estimates. They argue that elasticities are redundant, in that the data necessary for their determination could be used to determine total revenues directly. Thus managers could assess the effects of a change in price without knowledge of price elasticity. Although this is true, elasticity estimates are valuable, in that they provide a quick way of evaluating pricing policies. For example, if demand is known to be elastic, it is also known that a price increase will reduce total revenues.

**Key Concepts**

- Demand tends to be less elastic if:
  - There are few good substitutes available.
  - Expenditures on the good or service account for a small proportion of total expenditures.
  - Consumers have not yet had time to adjust fully to changes in price.
- A price *increase* will have one of the following effects, depending on the price elasticity of demand:
  - Total revenue will increase if demand is inelastic ($-1 < E_p < 0$).
  - Total revenue will decrease if demand is elastic ($E_p < -1$).
  - Total revenue will be unchanged if demand is unitary elastic ($E_p = -1$).

## Example How Much Tuition for College Students?

The board of trustees of a leading state university is faced with a critical financial prob-lem. At present tuition rates, the university is losing $7.5 million per year. The presiden of the university, a well-known biologist, urges that tuition be raised $750 over the pre sent $3,000 rate—a 25 percent increase. Based on the 10,000 students now attending th school, he projects that this increase would cover the $7.5 million shortfall in revenue

Student leaders protest that they cannot afford a tuition hike, but the president re sponds that the only alternative is to cut back significantly on programs and faculty. Th faculty supports the tuition increase as a means of preserving their jobs.

The students quickly realize that any appeal that involves compassion for th plight is likely to fall on deaf ears. Their only hope is to demonstrate that the tuition hi is not in the best interest of the university. What can they do?

**Solution** The university administration sees a tuition increase as a means of in-creasing total revenue. The board of trustees is likely to reject the administration's pro-posal only if the students can demonstrate that it would not achieve this purpose.

As part of a term paper, an economics major discovers a journal article that discusses the price elasticity of demand for a college education. The author of the article estimates that the elasticity for enrollment at state universities is $-1.3$ with respect to tuition changes. That is, a 1 percent increase in tuition would decrease enrollments by 1.3 per-cent. The data are current and the paper was written by a highly respected scholar.

Based on the elasticity estimate from the paper, the students calculate that the p posed tuition hike of 25 percent would decrease enrollment by 32.5 percent, or nearly 3 students. This would result in a decrease in total revenue from $30,000,000 at present ( $3,000 × 10,000) to about $25,000,000 after the tuition increase (i.e., $3,750 × 6,700).

Faced with this startling information, the board of trustees asks the university pr ident if the cost savings from fewer students would compensate for the revenue dr The president replies that most of the university's costs are independent of enrollm and, hence, there would not be a significant cost saving. By a unanimous vote, the bo of trustees rejects the tuition increase and orders the president to find some other v to meet the revenue deficiency.

## INCOME ELASTICITY

Income elasticities are used to measure the responsiveness of demand to changes in in-come. When other factors are held constant, the *income elasticity* of a good or service the percentage change in demand associated with a 1 percent change in income. Speci ically, the income elasticity of a good or service is defined as

$$E_I = \frac{\%\Delta Q}{\%\Delta I} = \frac{\Delta Q}{Q} \bigg/ \frac{\Delta I}{I} \qquad \text{(3-16)}$$

where $I$ denotes income. Rearranging terms, equation (3-16) can be written

$$E_I = \frac{\Delta Q}{\Delta I}\frac{I}{Q} \qquad \text{(3-17)}$$

As with price elasticity, income elasticity can be expressed in either arc or point terms. Arc income elasticity is used when relatively large changes in income are being considered and is defined as

$$E_I = \frac{Q_2 - Q_1}{I_2 - I_1} \frac{I_2 + I_1}{Q_2 + Q_1} \tag{3-18}$$

where $Q_1$ and $I_1$ represent the initial levels of demand and income, and $Q_2$ and $I_2$ are the values after a change in income.

For example, suppose that the demand for automobiles as a function of income per capita is given by the equation

$$Q = 50{,}000 + 5\,(I)$$

What is the income elasticity as per capita income increases from \$10,000 to \$11,000? Substituting $I_1 = \$10{,}000$ into the equation, quantity demanded is 100,000 cars. Similarly, at $I_2 = \$11{,}000$, quantity demanded is 105,000 automobiles. Thus

$$E_I = \frac{105{,}000 - 100{,}000}{11{,}000 - 10{,}000} \times \frac{11{,}000 + 10{,}000}{105{,}000 + 100{,}000} = 0.512$$

The interpretation of this result is that over the income range \$10,000 to \$11,000, each 1 percent increase in income causes about a five-tenths of 1 percent increase in quantity demanded.

If the change in income is small or if income elasticity at a particular income level needs to be determined, a point elasticity is appropriate. In this case, $\Delta Q/\Delta I$ is expressed as $dQ/dI$. Thus

$$E_I = \frac{dQ}{dI} \frac{I}{Q} \tag{3-19}$$

To illustrate, if $Q = 50{,}000 + 5I$ as before, each one-unit increase in income is associated with a five-unit increase in demand. Thus $dQ/dI = 5$. For $I = \$10{,}500$, demand is 102,500 units and the income elasticity is

$$E_I = 5 \times \frac{10{,}500}{102{,}500} = 0.512$$

Note that this value is identical to the arc elasticity between \$10,000 and \$11,000 of income. They are equal because demand is a linear function of income. Thus, the rate of change in quantity does not vary as income increases. Hence the average rate of change between \$10,000 and \$11,000 is the same as the instantaneous rate of change at \$10,500. However, for nonlinear demand equations, the arc and point elasticities at the midpoint of the arc are not necessarily equal.

## Inferior Goods, Necessities, and Luxuries

Income elasticities can be either negative or positive. When they are negative, an increase in income is associated with a decrease in the quantity demanded of the good or service. Cheap hot dogs might be an example. Those living on tight budgets may be unable to afford any other kind of meat. But as their incomes increase, they give up hot dogs and switch to other types of meat, such as roast beef and steak. Thus the increase in income causes a decrease in demand for hot dogs. Goods with negative income elasticities are defined as inferior goods.

Normal goods and services have positive income elasticities. They can be further classified by the magnitude of $E_I$. If $0 < E_I \leq 1$, the percentage change in demand is positive but less than or equal to the percentage change in income. Such goods and services are referred to as necessities. That is, demand is relatively unaffected by changes in income. A good example is bread, a basic food item eaten by even the poorest families. As a family becomes more affluent, it will consume more bread, but the increase is usually not proportionate to the increase in income.

Finally, luxuries are goods and services for which $E_I > 1$. This means that the change in demand is proportionately greater than the change in income. For example, if $E_I = 4$, a 1 percent increase in income would cause a 4 percent increase in demand. Jewelry is an example of a luxury good. As individuals become wealthier, they have more disposable income. Thus, purchases of necklaces, rings, and fine watches tend to represent a larger share of their incomes.

### Income Elasticity and Decision Making

The income elasticity for a firm's product is an important determinant of the firm's success at different stages of the business cycle. During periods of expansion, incomes are rising, and firms selling luxury items such as gourmet foods and exotic vacations will find that the demand for their products will increase at a rate that is faster than the rate of income growth. However, during a recession, demand may decrease rapidly. Conversely, sellers of necessities such as fuel and basic food items will not benefit as much during periods of economic prosperity, but will also find that their markets are somewhat recession-proof. That is, the change in demand will be less than that in the economy in general.

Knowledge of income elasticities can be useful in targeting marketing efforts. Consider a firm specializing in expensive men's colognes. Because such goods are luxuries those in high-income groups would be expected to be the prime customers. Thus the firm should concentrate its marketing efforts on media that reach the wealthier segments of the population. For example, advertising dollars should be spent on space in *Esquire* and the *New Yorker* rather than in the *National Enquirer* and *Wrestling Today*.

## *Case Study*

## Engel's Law and the Plight of the Farmer

In the nineteenth century, a German statistician, Ernst Engel, etched his name in economic history by proposing what has become known as *Engel's law.* Engel studied the consumption patterns of a large number of households and concluded that the percentage of income spent on food decreases as incomes increase. That is, he determined that food is a necessity. His finding has repeatedly been confirmed by later researchers. Examples of estimated income elasticities are shown in the following table. Note that only beef has an estimated elasticity greater than unity.

[illegible]

| *Food* | *Estimated Income Elasticity* |
|---|---|
| Beef | 1.05 |
| Chicken | 0.28 |
| Pork | 0.14 |
| Tomatoes | 0.24 |
| Potatoes | 0.15 |

*Sources:* D. B. Suits, "Agriculture," in *Structure of American Industry, ed. W. Adams, 8th ed.* (New York: Macmillan, 1990); and D. M. Shuffett, *The Demand and Price Structure for Selected Vegetables* (Washington, D.C.: U.S. Department of Agriculture, 1954, Technical Bulletin 1105).

One of the implications of Engel's law is that farmers may not prosper as much as those in other occupations during periods of economic prosperity. The reason is that if food expenditures do not keep pace with increases in gross domestic product, farm incomes may not increase as rapidly as incomes in general. However, this tendency has been partially offset by the rapid increase in farm productivity in recent years. In 1940, each U.S. farmer grew enough food to feed about 11 other people. Today, the typical farmworker produces enough to feed 80 people. ■

**Key Concepts**

- Income elasticity is the percentage change in demand per 1 percent change in income.
- Negative income elasticities denote inferior goods. Normal goods are those with positive income elasticities. If $0 < E_I \leq 1$, the product is defined as a necessity. For luxuries, $E_I > 1$.

## CROSS ELASTICITY

Demand is also influenced by prices of other goods and services. The responsiveness of quantity demanded to changes in price of other goods is measured by *cross elasticity*, which is defined as the percentage change in quantity demanded of one good caused by a 1 percent change in the price of some other good. That is,

$$E_C = \frac{\%\Delta Q_x}{\%\Delta P_y} \tag{3-20}$$

where $x$ and $y$ represent the goods or services being considered.

For large changes in the price of $y$, arc cross elasticity is used. The arc elasticity is computed as

$$E_C = \frac{Q_{x,2} - Q_{x,1}}{P_{y,2} - P_{y,1}} \frac{P_{y,2} + P_{y,1}}{Q_{x,2} + Q_{x,1}} \tag{3-21}$$

where the subscript 1 refers to the initial prices and quantities and 2 to the final values. Suppose that demand for $x$ in terms of the price of $y$ is given by

$$Q_x = 100 + 0.5P_y$$

If $P_y$ increases from \$50 to \$100, then, using the equation, it is determined that $Q_x$ increases from 125 to 150 units. Thus the cross price elasticity is

$$E_C = \frac{150 - 125}{100 - 50} \times \frac{100 + 50}{150 + 125} = 0.27$$

The interpretation is that a 1 percent increase in the price of $y$ causes a 0.27 percent increase in the quantity demanded of $x$.

Point cross elasticities are analogous to the point elasticities already discussed. For small changes in $P_y$,

$$E_C = \frac{dQ_x}{dP_y} \frac{P_y}{Q_x} \tag{3-22}$$

Based on the demand equation $Q_x = 100 + 0.5P_y$, the derivative, $dQ_x/dP_y = 0.5$. At $P_y = \$20$, quantity demanded is 110 units. Hence the point cross elasticity is

$$E_C = 0.5 \times \frac{20}{110} = 0.09$$

## Substitutes and Complements

Cross elasticities are used to classify the relationship between goods. If $E_C > 0$, an increase in the price of $y$ causes an increase in the quantity demanded of $x$, and the two products are said to be *substitutes*. That is, one product can be used in place of (substituted for) the other. Suppose that the price of $y$ increases. This means that the opportunity cost of $y$ in terms of $x$ has increased. The result is that consumers purchase less $y$ and more of the relatively cheaper good $x$. Beef and pork are examples of substitutes. An increase in the price of beef usually increases the demand for pork, and vice versa.

When $E_C < 0$, the goods or services involved are classified as *complements* (goods that are used together). Increases in the price of $y$ reduce the quantity demanded of that product. The diminished demand for $y$ causes a reduced demand for $x$. Bread and butter, cars and tires, and computers and computer software are examples of pairs of goods that are complements.

Cross elasticities are not always symmetrical. That is, a change in demand for good $x$ caused by a change in the price of good $y$ may not equal the change in demand for $y$ generated by a change in the price of $x$. Consider the case of margarine and butter. One study determined that a 1 percent increase in the price of butter caused a 0.81 percent increase in demand for margarine. However, a 1 percent increase in the price of margarine increased demand for butter by only 0.67 percent.[6] Although the two elasticities are different, note that they are both positive, indicating that butter and margarine are substitutes.

[6] H. Wold, *Demand Analysis* (New York: Wiley, 1953).

## Cross Elasticity and Decision Making

Many large corporations produce several related products. Gillette makes both razors and razor blades. Ford sells several competing makes of automobiles. Where a company's products are related, the pricing of one good can influence the demand for other products. Gillette will probably sell more razor blades if it lowers the price of its razors. In contrast, if the price of Fords is reduced, sales of Mercurys may decline. Information regarding cross elasticities can aid decision makers in assessing such impacts.

Cross elasticities are also useful in establishing boundaries between industries. Sometimes it is difficult to determine which products should be included in an industry. For example, should the manufacturing of cars and trucks be considered one industry or two? One way to answer such questions is to specify industries based on cross elasticities. This approach defines an industry as including firms whose products exhibit a high positive cross elasticity. Goods and services with negative or small cross elasticities are considered to belong to different industries. The definition of an industry might seem to be an unimportant matter, but the choice can have important implications. For example, the outcomes of antitrust cases alleging illegal monopolies are sometimes determined primarily by the industry definition accepted by the judges assigned to the case.

## Example Using Elasticities in Decision Making

The R. J. Smith Corporation is a publisher of romance novels—nothing exotic or erotic—just stories of common people falling in and out of love. The corporation hires an economist to determine the demand for its product. After months of hard work and submission of an exorbitant bill, the analyst tells the company that demand for the firm's novels ($Q_x$) is given by the following equation:

$$Q_x = 12{,}000 - 5{,}000P_x + 5I + 500P_c$$

where $P_x$ is the price charged for the R. J. Smith novels, $I$ is income per capita, and $P_c$ is the price of books from competing publishers.

Using this information, the company's managers want to

1. Determine what effect a price increase would have on total revenues
2. Evaluate how sale of the novels would change during a period of rising incomes
3. Assess the probable impact if competing publishers raise their prices

Assume that the initial values of $P_x$, $I$, and $P_c$ are \$5, \$10,000, and \$6, respectively.

**Solution**

1. The effect of a price increase can be assessed by computing the point price elasticity of demand. Substituting the initial values of $I$ and $P_c$ yields

$$Q_x = 12{,}000 + 5(10{,}000) + 500(6) - 5{,}000\,P_x$$

which is equivalent to

$$Q_x = 65{,}000 - 5{,}000P_x$$

Note that $dQ_x/dP_x = -5{,}000$. At $P_x = \$5$, quantity demanded is 40,000 books. Using these data, the point price elasticity is computed to be

$$E_p = -5{,}000 \times \frac{5}{40{,}000} = -0.625$$

Because demand is inelastic, raising the price of the novels would increase total revenue.

2. The income elasticity determines whether a product is a necessity or a luxury. It has already been determined that the initial quantity demanded at the given values of the price and income variables is 40,000. From the demand equation the derivative, $dQ_x/dI = 5$. Thus the income elasticity is

$$E_I = 5 \times \frac{10{,}000}{40{,}000} = 1.25$$

Because $E_I > 1$, the novels are a luxury good. Thus as incomes increase, sales should increase more than proportionately.

3. The demand equation implies that $dQ_x/dP_c = 500$. Thus it is known that $E_C$ is positive, meaning that Smith's romance novels and books from competing publishers are viewed by consumers as substitutes. Computing $E_C$ yields

$$E_C = 500 \times \frac{6.00}{40{,}000} = 0.075$$

Hence, a 1 percent increase in the price of other books results in a 0.075 percent increase in demand for R. J. Smith's romance novels.

---

**Key Concepts**

- Cross elasticity [illegible] change in quantity demanded of good x per 1 percent change [illegible] good y.
- Cross elasticities [illegible] negative for complements.

---

## SUMMARY

Demand refers to the number of units of a good or service that consumers are willing and able to buy at each price during a specified interval of time. Market demand is the sum of all individual demands. A movement along the demand curve is caused by changes in price and is referred to as a change in quantity demanded. A change in demand is represented by a shift of the demand curve. Changes in demand can be caused by changes in tastes and preferences, income, and prices of other goods and services.

Marginal revenue is the change in total revenue per one-unit change in demand. Total revenue is increasing when marginal revenue is positive. Marginal revenue is zero at the point of maximum total revenue and total revenue is declining when marginal revenue is negative. For linear demand curves, the absolute value of the slope of the marginal revenue curve is twice that of the demand curve.

Elasticities measure the responsiveness of demand to various factors. Price elasticity is defined as the percentage change in quantity demanded per 1 percent change in price. Arc price elasticity is used to assess the impact of discrete changes in price. Point price elasticity is appropriate for very small price changes.

Demand is said to be elastic where $E_p < -1$ and inelastic for $-1 < E_p \leq 0$. Demand is unitary elastic if $E_p = -1$. Marginal revenue is positive where demand is elastic, zero for unitary elasticity, and negative for inelastic demand. Demand tends to be more elas tic when (1) there are many substitutes for a good or service, (2) a substantial proportion of total income is spent on the product, and (3) the time period is longer.

Income elasticity is the percentage change in quantity demanded per 1 percent change in income. Inferior goods have negative income elasticities, while normal goods are those with positive income elasticities. For necessities, $0 < E_I \leq 1$. Luxuries are goods and services with income elasticities greater than 1.

Cross elasticity is defined as the percentage change in quantity demanded of one good per 1 percent change in the price of some other good. Cross elasticities are positive for substitutes and negative for complements. Goods with high positive cross elasticities are considered to be in the same market.

## Discussion Questions

3-1. Suppose that the possession of marijuana was legalized in all states. What would happen to the demand for marijuana? Explain.

3-2. Why is the market demand curve usually less elastic than demand curves faced by the individual firms in that market?

3-3. The income elasticity for soft drinks is estimated to be 1.00. How could an executive of a bottling firm use this information to forecast sales?

3-4. The profit-maximizing price will never be set where demand is inelastic. True or false? Explain.

3-5. If demand is unitary elastic, what action could a manager take to increase total revenue? Explain.

3-6. Why would demand for natural gas be more inelastic in the short run than in the long run?

3-7. Which of the following pairs of goods, are substitutes and which are complements? Explain.
   a. Insulation and heating oil
   b. Hot dogs and mustard
   c. Television sets and videocassette recorders
   d. Rice and potatoes

3-8. In a world with just two goods where all income is spent on the two goods, both goods cannot be inferior. True or false? Explain.

3-9. In 1988, tuition at Commonwealth College was $4,000 per year and enrollment was 5,000 students. By 1998, tuition had increased to $8,000, but enrollment had increased to 5,500 students. Does this change imply an upward-sloping demand curve? Explain.

## Problems

3-1. It is known that quantity demanded decreases by two units for each $1 increase in price. At a price of $5, quantity demanded is ten units.
   a. What will be the quantity demanded if price is zero?
   b. Write an equation for quantity demanded as a function of price.
   c. Write an equation that expresses price as a function of quantity.
   d. Write an equation for total revenue.

3-2. ~~A market consists~~ of three people, $A$, $B$, and $C$, whose individual demand equations are as follows:

$$A: \quad P = 35 - 0.5Q_A$$
$$B: \quad P = 50 - 0.25Q_B$$
$$C: \quad P = 40 - 2.00Q_C$$

The industry supply equation is given by $Q_S = 40 + 3.5P$.

a. Determine the equilibrium price and quantity.

b. Determine the amount that will be purchased by each individual.

3-3. A market consists of two individuals. Their demand equations are $Q_1 = 16 - 4P$ and $Q_2 = 20 - 2P$, respectively.

a. What is the market demand equation? $Q_1 + Q_2$

b. At a price of \$2, what is the point price elasticity for each person and for the market?

3-4. The demand equation faced by DuMont Electronics for its personal computers is given by $P = 10{,}000 - 4Q$.

a. Write the marginal revenue equation.

b. At what price and quantity will marginal revenue be zero?

c. At what price and quantity will total revenue be maximized?

d. If price is increased from \$6,000 to \$7,000, what will the effect be on total revenue? What does this imply about price elasticity?

3-5. The demand for shirts produced by a Canadian manufacturer has been estimated to be $P = 30 - Q/200$.

a. Compute the point elasticity at $P = \$10$; at $P = \$15$.

b. How does the point elasticity vary with the price?

3-6. A manager believes that the demand for her product is given by the equation $P = 50 - Q/100$.

a. What is the arc elasticity of demand as price decreases from \$12 to \$10?

b. What is the arc elasticity of demand as price increases from \$10 to \$12?

3-7. For each of the following equations, determine whether demand is elastic, inelastic, or unitary elastic at the given price.

a. $Q = 100 - 4P$ and $P = \$20$.

b. $Q = 1500 - 20P$ and $P = \$5$.

c. $P = 50 - 0.1Q$ and $P = \$20$.

3-8. Sailright Inc. manufactures and sells sailboards. Management believes that the price elasticity of demand is $-3.0$. Currently, boards are priced at \$500 and the quantity demanded is 10,000 per year.

a. If the price is increased to \$600, how many sailboards will the company be able to sell each year?

b. How much will total revenue change as a result of the price increase?

3-9. Demand for a softback managerial economics text is given by $Q = 20{,}000 - 300P$. The book is initially priced at \$30.

a. Compute the point price elasticity of demand at $P = \$30$.

b. If the objective is to increase total revenue, should the price be increased or decreased? Explain.

c. Compute the arc price elasticity for a price decrease from \$30 to \$20.

d. Compute the arc price elasticity for a price decrease from \$20 to \$15.

3-10. Write a demand equation for which the price elasticity of demand is zero for all prices.

3-11. A consultant estimates the price–quantity relationship for New World Pizza to be $P = 50 - 5Q$.

a. At what output rate is demand unitary elastic?

b. Over what range of output is demand elastic?

c. At the current price, eight units are demanded each period. If the objective is to increase total revenue, should the price be increased or decreased? Explain.

3-12. The price elasticity for rice is estimated to be −0.4 and the income elasticity is 0.8. At a price of $0.40 per pound and a per capita income of $20,000, the demand for rice is 50 million tons per year.

a. Is rice an inferior good, a necessity, or a luxury? Explain.

b. If per capita income increases to $20,500, what will be the quantity demanded of rice?

c. If the price of rice increases to $0.41 per pound and income per capita remains at $20,000, what will be the quantity demanded?

3-13. Acme Tobacco is currently selling 5,000 pounds of pipe tobacco per year. Due to competitive pressures, the average price of a pipe declines from $15 to $12. As a result, the demand for Acme pipe tobacco increases to 6,000 pounds per year.

a. What is the cross elasticity of demand for pipes and pipe tobacco?

b. Assuming that the cross elasticity does not change, at what price of pipes would the demand for pipe tobacco be 3,000 pounds per year? Use $15 as the initial price of a pipe.

3-14. The McNight company is a major producer of steel. Management estimates that the demand for the company's steel is given by the equation

$$Q_s = 5{,}000 - 1{,}000P_s + 0.1I + 100P_a$$

where $Q_s$ is steel demand in thousands of tons per year, $P_s$ is the price of steel in dollars per pound, $I$ is income per capita, and $P_a$ is the price of aluminum in dollars per pound. Initially, the price of steel is $1 per pound, income per capita is $20,000, and the price of aluminum is $0.80 per pound.

a. How much steel will be demanded at the initial prices and income?

b. What is the point income elasticity at the initial values?

c. What is the point cross elasticity between steel and aluminum? Are steel and aluminum substitutes or complements?

d. If the objective is to maintain the quantity of steel demanded as computed in part (a), what reduction in steel prices will be necessary to compensate for a $0.20 reduction in the price of aluminum?

3-15. The price of oil is $30 per barrel and the price elasticity is constant and equal to −0.5. An oil embargo reduces the quantity available by 20 percent. Use the arc elasticity formula to calculate the percentage increase in the price of oil.

3-16. The arc advertising elasticity is 1.5 as advertising expenditure increases from $10 to $12 million. If demand is 50 at an advertising expenditure of $12 million, what will demand be at an advertising expenditure of $10 million?

3-17. The demand equation is estimated to be $50 - 3P + 2P_o$, where $P_o$ is the price of some other good. Assume the average value of $P$ is $3 and the average value of $P_o$ is $6.

a. What is the price elasticity at the average values of $P$ and $P_o$? How should the price of the good be changed to increase total revenues?

b. What is the cross elasticity at the average values of $P$ and $P_o$? What is the relationship between the two goods?
c. If the equation is correctly estimated, is the good inferior, a necessity, or a luxury? Explain.

## Problems Requiring Calculus

3-18. The Inquiry Club at Jefferson University has compiled a book that exposes the private lives of many of the professors on campus. Economics majors in the club estimate that total revenue from sales of the book is given by the equation

$$TR = 120Q - 0.1Q^3$$

a. Over what output range is demand elastic?
b. Initially, the price is set at \$71.60. To maximize total revenue, should the price be increased or decreased? Explain.

3-19. The demand equation for a product is given by

$$P = 30 - 0.1Q^2$$

a. Write an equation for the point elasticity as a function of quantity.
b. At what price is demand unitary elastic?

3-20. The demand equation for a product is given by

$$Q = \frac{20I}{P}$$

where $I$ is income and $P$ is price.
a. Write an equation for the point price elasticity. For what values of $I$ and $P$ is demand unitary elastic? Explain.
b. Write an equation for the point income elasticity. For what values of $I$ and $P$ is the good a necessity? Explain.

3-21. Given the demand equation, $Q = 12{,}000 - 10P^2$.
a. For this equation, write the expression for the point price elasticity of demand as a function of $P$.
b. Over what range of prices is the demand inelastic?

3-22. Consider the demand equation, $Q_x = 150 - P_xP_o$ where the subscripts $x$ and $o$ refer to two different goods.
a. For this equation, write the expression for the point price and cross elasticities of demand as a function of $P_o$ and $P_x$.
b. What is the relationship between the price and cross elasticities?

# Appendix

## *Behind the Demand Curve: The Theory of Consumer Choice*

The discussion of demand in chapter 3 began by assuming that the relationship between price and quantity demanded was known. The purpose of this appendix is to explain the derivation of the demand function and to develop a theory of consumer choice. This theory provides an understanding of the consumer decision-making process. It also serves as a guide for predicting consumer behavior under various conditions.

## CONSUMER PREFERENCES

Individuals make choices based on their personal tastes and preferences. Those who enjoy skiing may spend a large percentage of their income on ski equipment, whereas those who are inclined to less strenuous activities may use their income to buy books or CDs. Tastes and preferences are shaped by many factors. Some of the determinants are family environment, physical condition, age, sex, education, religion, and location. In the analysis that follows, tastes and preferences will be viewed as a given, and the discussion will focus on how those tastes and preferences are transformed into consumption decisions.

In the well-established tradition of economics, four basic assumptions are made in developing the model of consumer choice. First, it is assumed that individuals can rank their preferences for alternative bundles of goods and services. Consider a world in which only two goods are available: deodorant and mouthwash. Suppose that a consumer is confronted with the following combinations of those two goods:

| *A* | *B* | *C* |
|---|---|---|
| 5 cans deodorant<br>2 bottles mouthwash | 2 cans deodorant<br>6 bottles mouthwash | 4 cans deodorant<br>4 bottles mouthwash |

The ability to rank means that the individual can assess the relative amount of satisfaction that would result from each bundle of goods. For example, suppose that *B* is considered the most desirable bundle, and that *C* and *A* are viewed as providing equal but lesser amounts of satisfaction than *B*. Using the terminology of the theory of consumer choice, it is said that *B* is preferred to both *C* and *A* and that the consumer is indifferent between *C* and *A*.

The second assumption is *nonsatiation*. This means that individuals consider themselves better off if they have more of a good or service than if they have less. Consider a bundle *D* that consists of

| D |
| --- |
| 3 cans deodorant |
| 7 bottles mouthwash |

The nonsatiation assumption implies that this bundle would be preferred to bundle *B* because it includes more deodorant *and* more mouthwash.

*Transitivity* is the third assumption. It can be thought of as requiring that preferences be consistent. Transitivity states that if bundle *D* is preferred to bundle *B* and if *B* is preferred to both *A* and *C*, bundle *D* must be preferred to bundles *A* and *C*.

Finally, it is assumed that in order to get additional units of one good, consumers are willing to give up successively fewer units of other goods. For example, a consumer may be willing to forgo the purchase of five cans of deodorant to obtain the first bottle of mouthwash. However, if the person already has three bottles of mouthwash, the value of another bottle in terms of deodorant is likely to be less than five cans.

**Key Concepts**

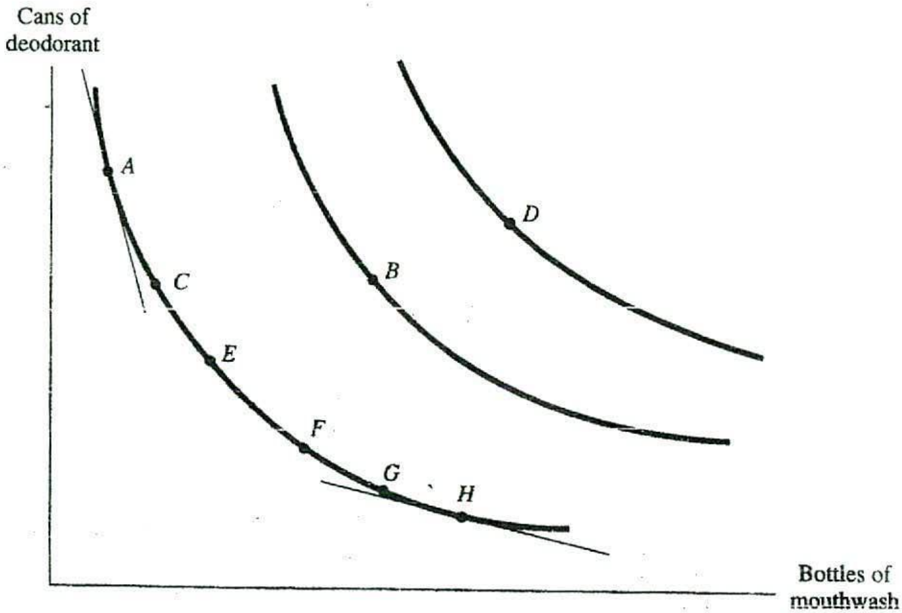
- Four assumptions form the basis for the theory of consumer choice. They are:
  1. Individuals can rank their preferences.
  2. Nonsatiation—people prefer more to less.
  3. Transitivity—rankings are consistent.
  4. Individuals are willing to give up successively smaller amounts of one good in order to get additional units of other goods.

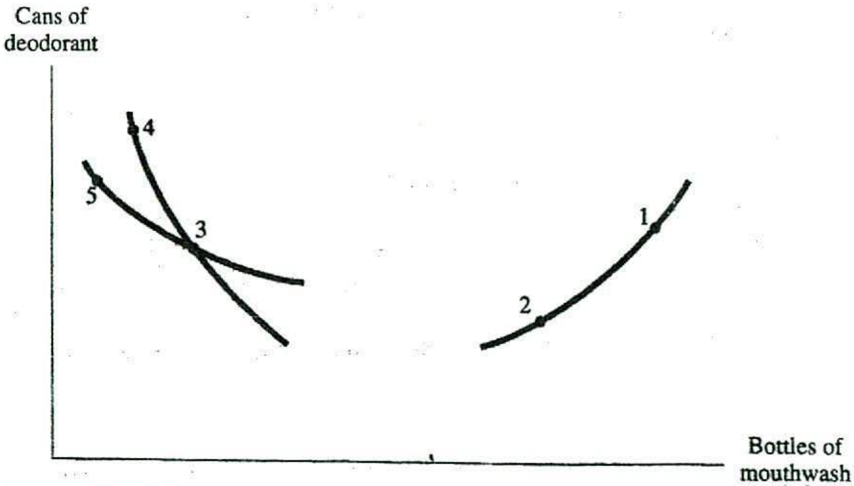
## INDIFFERENCE CURVES

Recall that bundles *A* and *C* were viewed as being equally desirable. That is, an individual would be indifferent to having *A* rather than *C*. If asked to select one over the other, there would be no basis for choice. Now suppose that other bundles, designated as *E*, *F*, *G*, and *H*, are also considered equivalent to *A* and *C*. If these bundles are plotted on a graph as shown in Figure 3A.1, the points can be joined to form an *indifference curve* that represents all bundles of goods that provide an individual with equal levels of satisfaction.

Note that Figure 3A.1 shows several indifference curves. All points on the curve through *B* are considered equivalent to that bundle. Because *B* is preferred to *A*, the assumption of transitivity guarantees that all the points on the indifference curve passing through *B* are preferred to all the points on the curve associated with *A*. Similarly, since *D* provides more satisfaction than *B*, all points on the curve passing through *D* are preferred to those on the curves passing through *B* and *C*. Because of nonsatiation, higher indifference curves denote increased levels of satisfaction.

The four assumptions stated earlier determine the basic characteristics of indifference curves. The assumption that individuals are capable of ranking their preferences implies that indifference curves exist. The assumption of nonsatiation assures that the curves will have a negative slope. This is easily shown by considering a curve with posi-




tive slope, such as the curve passing through point 1 in Figure 3A.2. Pick any other point on the curve, such as 2. Note that point 1 denotes a bundle with more of both mouthwash and deodorant than point 2. But because of the nonsatiation assumption, having more of both goods implies that 1 is preferred to 2. Thus, the two points cannot be on the same indifference curve. Hence, indifference curves must be downward sloping.

Transitivity and nonsatiation guarantee that two indifference curves will not intersect. This can also be seen from Figure 3A.2, which shows two indifference curves crossing at point 3. Consider points 4 and 5. The bundle denoted by 4 has more of both goods than 5 and hence must be preferred to 5. Because 3 and 5 are on the same indifference curve, transitivity requires that 4 be preferred to 3. But this is not true; 3 and 4 are on the same indifference curve. Thus the assumption of transitivity has been violated because preferences are not consistent. The assumption of transitivity is always violated when indifference curves intersect.

The assumption that consumers will be willing to give up successively fewer units of one good in order to get additional units of another good assures that indifference curves will have the convex shape shown in Figure 3A.1. The slope at any point on an indifference curve is the tangent to the line at that point. This slope represents the number of cans of deodorant that will be given up to get one more bottle of mouthwash, while still remaining on the same indifference curve. That is, it is the rate at which the individual is willing to trade deodorant for mouthwash. This trade-off is referred to as the *marginal rate of substitution* (*MRS*). For example, in Figure 3A.1, if the slope of the tangent to the indifference curve at point $A$ is $-8$, the individual will be willing to substitute eight cans of deodorant for one bottle of mouthwash.

To satisfy the fourth assumption, the absolute value of the MRS must decrease when moving down an indifference curve. Indifference curves that are convex to the origin have this property. Note that the absolute value of the slope of the curve in Figure 3A.1 decreases from point $A$ to point $H$. If the slope of the tangent at point $H$ is $-1/8$, the individual is now willing to give up eight bottles of mouthwash to get a single can of deodorant. An indifference curve must be convex in order to depict this declining marginal rate of substitution.

**Key Concepts**

- Indifference curves are downward sloping, convex, and do not intersect. Higher indifference curves reflect greater levels of satisfaction.
- The slope of an indifference curve is called the marginal rate of substitution (*MRS*). It denotes the rate at which an individual is willing to trade two goods or services.
- Convex indifference curves depict a declining marginal rate of substitution.

## BUDGET CONSTRAINTS

Wants reflect individual tastes and preferences. But actual purchases are strongly influenced by income and prices. Let $I$ be income and $P_D$ and $P_M$ be the prices of deodorant and mouthwash, respectively. In a two-good world, possible purchases are defined by the following expression:

$$I \geq P_D Q_D + P_M Q_M \tag{3A-1}$$

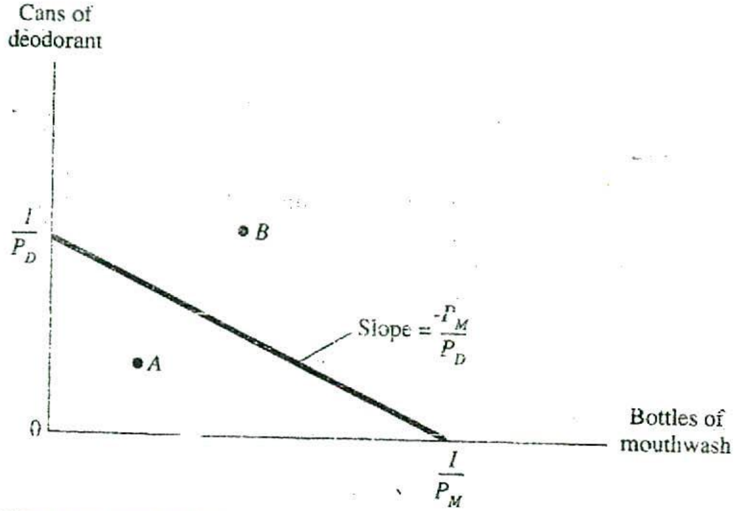



**FIGURE 3A.3 Budget Constraint**

where $Q_D$ and $Q_M$ are the quantities of deodorant and mouthwash purchased. The interpretation of equation (3A-1) is straightforward. It states that the sum of money spent on deodorant plus the sum spent on mouthwash must be less than or equal to the total income available.

If equation (3A-1) is treated as an equality, it denotes the possible bundles of deodorant and mouthwash that can be purchased if all the income is spent. Solving this equation for $Q_D$ yields

$$Q_D = \frac{I}{P_D} - \frac{P_M}{P_D} Q_M \qquad \text{(3A-2)}$$

which expresses possible purchases of deodorant in terms of income, prices, and the quantity of mouthwash purchased. Equation (3A-2) can be graphed as shown in Figure 3A.3.

The line in Figure 3A.3 is called a *budget constraint.* All the points to the left of the curve, such as point *A*, represent quantities of deodorant and mouthwash that can be purchased using less than all the available income. Points to the right of the budget constraint, such as point *B*, are bundles that cannot be purchased with the available income.

Note that the vertical intercept of the budget constraints is $I/P_D$. It represents the number of cans of deodorant that could be purchased if all income was spent on that commodity. The horizontal intercept is $I/P_M$ and has a similar interpretation. The slope of the budget constraint is $-P_MP_D$. It represents the rate at which one good can be substituted for another in the marketplace. For example, if $P_M/P_D = 1/2$, two bottles of mouthwash must be given up to get one more can of deodorant.

## UTILITY MAXIMIZATION

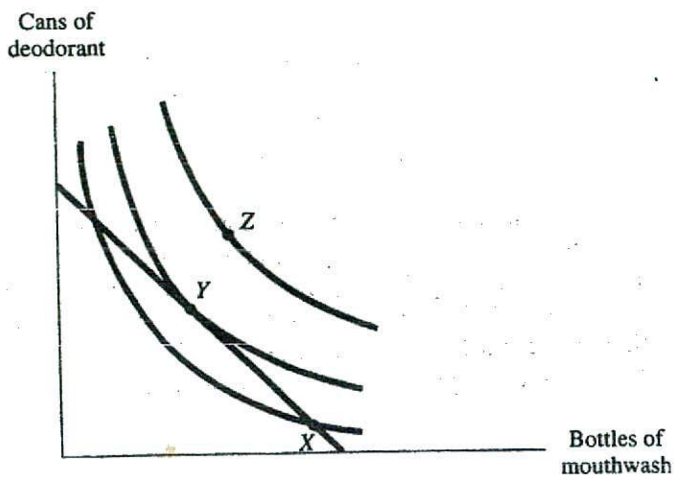
It is assumed that individuals strive to achieve the most satisfaction possible from their purchase choices. This objective is often referred to as *utility maximization.* It involves consideration of both indifference curves and the budget constraint. Suppose that a consumer is contemplating the purchase of bundle *X* as shown in Figure 3A.4. Point *X* lies on the budget constraint, indicating that it is an affordable bundle. But *X* is not the bundle that will maximize the individual's utility. Note that *Y* is also on the budget constraint but is on a higher indifference curve. Thus *Y* is preferred to *X*.

Point *Y* lies at the point of tangency between the budget constraint and the indifference curve. That is, point *Y* is the only point on the curve that touches the budget constraint. Note that a point such as *Z* on any higher indifference curve is above the budget constraint and hence is not affordable. Thus the bundle represented by point *Y* is the utility-maximizing point.

Given the individual's tastes, preferences, and income, and the prices of the two goods, there is no other point that will provide the same level of utility or satisfaction as point *Y*. Because *Y* is the point of tangency between the indifference curve and the budget constraint, the slope of the two lines is equal at that point. Remember that the slope of an indifference curve is the marginal rate of substitution, and the slope of the budget constraint is the negative of the price ratio, $-P_M/P_D$. Hence the utility maximizing bundle is the point where

$$-\frac{P_M}{P_D} = MRS \qquad \text{(3A-3)}$$

Equation (3A-3) is easily interpreted. The price ratio represents the rate at which the market requires consumers to substitute the two goods. The *MRS* is the rate at which the individual desires to substitute the goods. Utility maximization occurs where the rate at which the consumer wants to substitute is just equal to the rate at which he or she must substitute. If these two rates are not equal, purchases can be rearranged in such a way as to increase satisfaction or utility.




[redacted]

### Key Concepts

- The utility-maximizing point occurs where the highest indifference curve is tangent to the budget constraint.
- At the utility-maximizing point, the rate that products must be traded in the market is just equal to the rate at which the individual is willing to substitute one good for another.

## CONSUMER CHOICE AND THE DEMAND CURVE

The price–quantity combinations shown on an individual demand curve are the result of utility-maximizing decisions at various prices. This can easily be seen by considering the demand for mouthwash. Suppose that tastes and preferences, income, and the prices of deodorant and mouthwash generate indifference curves and budget constraints as shown in Figure 3A.5. If mouthwash costs \$1 per bottle, the utility-maximizing bundle will be point *Y*, which denotes five bottles of this good. That is, at $P = \$1$, quantity demanded is five bottles. Thus *Y* corresponds to one point on the demand curve for mouthwash.

If the price of mouthwash changes, *Y* will no longer be the utility-maximizing bundle. Suppose that $P_M$ increases to \$2. This change causes the budget constraint to rotate in a clockwise manner, as shown in Figure 3A.5. Note that the vertical intercept is the same as before ($I/P_D$). But the slope of the budget constraint is $-P_M/P_D$. Thus the increase in $P_M$ causes the line to rotate clockwise.

With the new budget constraint, the point *Y* is no longer affordable. Hence the consumer must select a new utility-maximizing bundle. As before, this point is where the highest indifference curve is tangent to the budget constraint. In Figure 3A.5 the utility-maximizing point is *Z*. Note that *Z* is a bundle with less mouthwash (3 bottles compared to 5) than point *Y*. The higher price of mouthwash reduced purchasing power and increased the opportunity cost of mouthwash relative to deodorant.

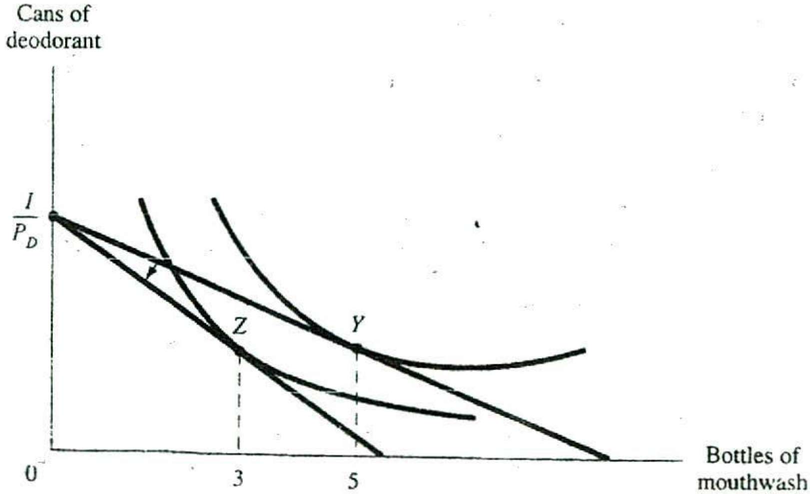



**FIGURE 3A.5 Price Changes and Utility Maximization**

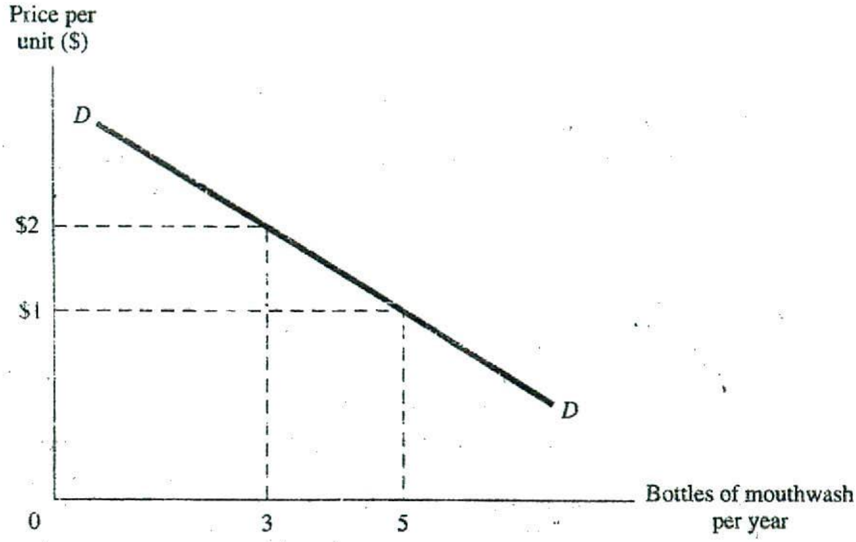



FIGURE 3A.6 Demand Curve for Mouthwash

Point $Z$ represents another price–quantity point on the demand curve. It is the quantity demanded when $P_M = \$2$. As shown by Figure 3A.6, the individual demand curve is generated by plotting these quantities as a function of price. Additional points are obtained by changing the price of mouthwash and determining the utility-maximizing quantity of the product.

## CONSUMER CHOICE AND CHANGES IN DEMAND

In the preceding section, the theory of consumer choice was used to show how changes in the quantity demanded (movements along the demand curve) occur. The theory can also be used to explain changes in demand (i.e., shifts of the demand curve). Such shifts result from a change in tastes and preferences, a change in income, or a change in the price of other goods. Consider the impact of an increase in income. When income increases, the budget constraint shifts outward, as shown in Figure 3A.7. Because the slope of the constraint is determined by prices and not by income, the new budget constraint is parallel to the initial budget constraint.

Initially, the utility-maximizing point was at $Y$ where $I_1$ is tangent to the budget constraint, but the extra income allows movement to a higher indifference curve. The utility-maximizing point is now $Z$, where $I_2$ is tangent to the new budget constraint. Note that $Z$ involves more mouthwash than did $Y$. At the same price of mouthwash, there is a greater demand for mouthwash than before. The result of this income change would appear as a rightward shift of the demand curve—more mouthwash would be demanded at each price.

The effect on demand of changes in tastes and preferences is illustrated by Figure 3A.8. Initially, let the utility-maximizing point be $Y$, as shown in the "before" panel. Now, suppose that concern about environmental impacts of a deodorant ingredient causes the consumer to view deodorant less favorably. This is indicated by a shift in the

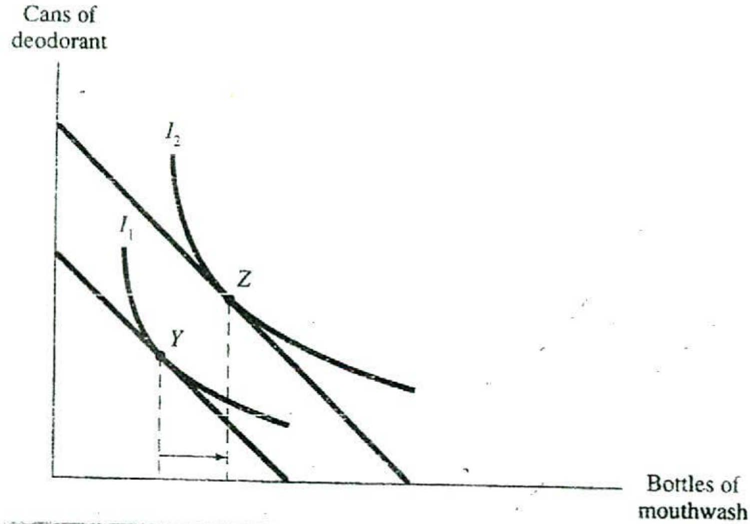



FIGURE 3A.7 Income Changes and Utility Maximization

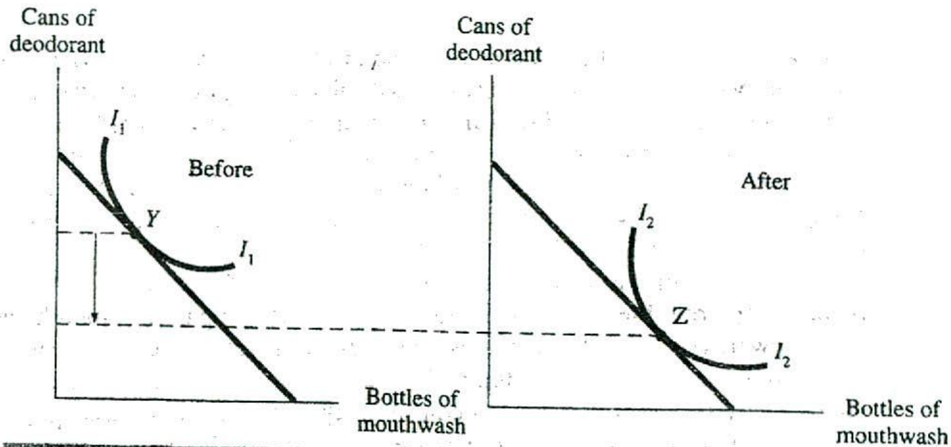



FIGURE 3A.8 Changes in Tastes and Preferences and Utility Maximization

indifference curve from $I_1$ to $I_2$ and a new utility-maximizing point, *Z*, as shown in the "after" panel. Note that at the new utility-maximizing point, less deodorant is selected than before the change in preferences. This would be represented on the demand curve as a leftward shift. That is, less deodorant would be demanded at any price.

In a similar way, the theory of consumer choice can be used to show how changes in the prices of other goods cause the demand curve to shift. Point *Y* in Figure 3A.9 is the utility-maximizing bundle for given income, tastes and preferences, and prices of deodorant and mouthwash.

Now, suppose that the price of mouthwash decreases. This is depicted in Figure 3A.9 by a counterclockwise rotation of the budget constraint. The new utility-maximizing bundle is point *Z*, where $I_2$ is tangent to the new budget constraint. Note that this point spec-

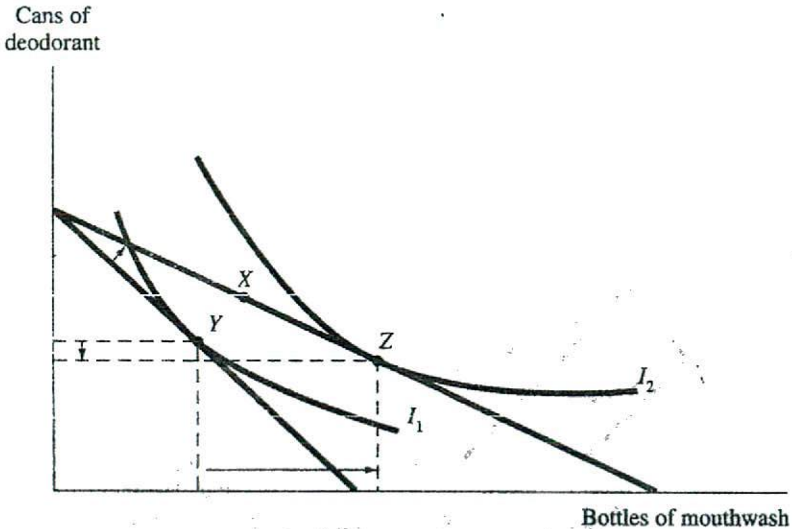



**FIGURE 5A.4 Changes in the Price of Other Goods and Utility Maximization**

ifies more mouthwash than was selected before the price reduction. This is because the opportunity cost of buying mouthwash has declined. The price change would be depicted as a movement from one point to another along the demand curve for mouthwash.

Note also that point *Z* involves less deodorant than did point *Y*. That is, the decrease in the price of mouthwash resulted in a decrease in the demand for deodorant. This change causes a leftward shift in the demand curve for deodorant—less will be demanded at each price.

The impact of a price change of one good on the demand for the other good depends on the relationship between the two products. Apparently, mouthwash and deodorant were considered by the consumer to be substitutes. Thus a decrease in the price of mouthwash decreased the demand for deodorant. If the goods had been complements, analysis using the theory of consumer choice would have predicted an increase in the demand for deodorant. In this case, a decrease in the price of mouthwash would have resulted in a utility-maximizing point such as *X*, which specifies more deodorant than before the price change.

## Key Concepts

- Changes in income are shown by a parallel shift of the budget constraint. A change in the price of one good causes the budget constraint to rotate.
- At different prices, the quantities demanded shown by the demand curve correspond to the points of tangency between budget constraints and indifference curves.
- Shifts of the demand curve reflect changes in the budget constraint or indifference curves caused by changes in income, prices of other goods, or tastes and preferences.

## Discussion Questions

3A-1. "The price ratio is the rate at which individuals must substitute goods in the marketplace." Explain this statement.

3A-2. How do changes in income affect the slope of the budget constraint? Explain.

3A-3. Suppose that bundle *A* is preferred to *B*, *C* is preferred to *D*, and *D* is preferred to *B*. What can be said about *C* in relation to *B*? What about *C* in relation to *A*? Are any of the assumptions of the theory of consumer choice violated by these preferences?

## Problems

3A-1. Consider a world in which there are only two goods. An individual has an income of $30,000, the price of deodorant is $3 per can, and the price of mouthwash is $4 per bottle.

a. Expressing deodorant as the dependent variable, write the equation for the budget constraint.

b. What is the slope of the budget constraint?

c. If the price of mouthwash increases to $5, write the new equation for the budget constraint.

3A-2. Assume that the budget constraint is given by the equation $Q_1 = 1{,}000 - 5Q_2$, where $Q_1$ and $Q_2$ represent quantities of two goods. Normally, indifference curves are convex to the origin, but assume in this case that they are linear with a constant slope of $-2$.

a. Graph the budget constraint (with $Q_1$ on the vertical axis).

b. Draw in a set of indifference curves and label the utility-maximizing point.

c. Where would the utility-maximizing point have been if the indifference curves had a constant slope of $-6$?

# CHAPTER 4

# Regression Techniques and Demand Estimation

## PREVIEW

In the previous chapter, demand analysis was introduced as a tool for managerial decision making. For example, it was shown that a knowledge of price and cross elasticities can assist managers in pricing and that income elasticities provide useful insights into how demand for a product will respond to different macroeconomic conditions.

In chapter 3, it was assumed that these elasticities were known or that the data were already available to allow them to be easily computed. Unfortunately, this is not usually the case. For many business applications, the manager who desires information about elasticities must develop a data set and use statistical methods to estimate a demand equation from which the elasticities can then be calculated. This estimated equation can then also be used to predict demand for the product, based on assumptions about prices, income, and other factors.

In this chapter, the basic techniques of demand estimation are introduced. The first major section considers regression analysis, which is a statistical method for fitting an equation to a data set. Regression analysis is used for demand estimation in this chapter and is also the technique used to estimate production and cost equations in later chapters. The second section describes the four basic steps involved in estimating a demand equation. Finally, three potential difficulties associated with regression analysis are discussed—omitted variables, the identification problem, and multicollinearity.

## REGRESSION TECHNIQUES

Consider the simple demand equation $Q_d = B + aP$. The law of demand implies that the coefficient $a$ should be negative, indicating that less of the product is demanded at higher prices. However, in making pricing decisions, it may not be sufficient to know that quantity demanded and price are inversely related. An estimate of the numerical value of $a$ and also of the coefficient $B$ may be required for decision making. Similarly, to fully understand the correspondence between inputs and outputs in production functions and between costs and output in cost functions, it is often necessary to quantify relationships between the variables. The most widely used technique in economics and many other sciences for estimating these relationships is the least-squares regression method. The basic elements of this technique are developed in this section.

Although the focus of this chapter is on estimation of demand, the example used in this section to introduce regression analysis is based on cost and output data. The reason for selecting a cost function is that it is somewhat easier to explain and to understand in a regression context than are price–quantity relationships. After the basics of least-squares analysis have been presented, the remainder of the chapter considers how the method is used to estimate demand functions.

### Estimating Coefficients

Consider a firm with a fixed capital stock that has been rented under a long-term lease for $100 per production period. The other input in the firm's production process is labor, which can be increased or decreased quickly depending on the firm's needs. In this case, the cost of the capital input ($100) is fixed and the cost of labor is variable. The manager of the firm wants to know the relationship between output and cost, that is, the firm's total cost function. This would allow the manager to predict the cost of any specified rate of output for the next production period.

Specifically, the manager is interested in estimating the coefficients $a$ and $b$ of the function

$$Y = a + bX$$

where the dependent variable $Y$ is total cost and the independent variable $X$ is total output. If this function is plotted on a graph, the parameter $a$ would be the vertical intercept (i.e., the point where the function intersects the vertical axis) and $b$ would be the slope of the function. Recall that the slope of a total function is the marginal function. As $Y = a + bX$ is the total cost function, the slope, $b$, is marginal cost or the change in total cost per unit change in output.

Assume that data on cost and output have been collected for each of seven production periods and are reported in Table 4.1. Note that there is a cost of \$100 associated with an output rate of zero. This represents the fixed cost of the capital input, which must be paid regardless of the rate of output. These data are shown as points in Figure 4.1. They suggest a definite upward trend, but they do not trace out a straight line. The problem is to determine the line that best represents the overall relationship between $Y$ and $X$. One approach would simply be to "eyeball" a line through these data in a way that the data points were about equally spaced on both sides of the line. The coefficient $a$ would be found by extending that line to the vertical axis and reading the $Y$-coordinate at that point. The slope, $b$, would be found by taking any two points on the line, $\{X_1, Y_1\}$ and $\{X_2, Y_2,\}$ and using the slope formula

$$b = \frac{Y_2 - Y_1}{X_2 - X_1}$$

Although this approach could be used, the method is quite imprecise and can be employed only when there is just one independent variable. What if production cost depends on both the rate of output and the size of the plant? To plot the data for these three variables (total cost, output, and plant size) would require a three-dimensional diagram; it would be nearly impossible to eyeball the relationship in this case. The addition of another independent variable, say average skill levels of the employees, would place the data set in the fourth dimension, where any graphic approach is hopeless.

| *Production Period* | *Total Cost* $(Y_i)$ | *Total Output* $(X_i)$ |
|---|---|---|
| 1 | \$100 | 0 |
| 2 | 150 | 5 |
| 3 | 160 | 8 |
| 4 | 240 | 10 |
| 5 | 230 | 15 |
| 6 | 370 | 23 |
| 7 | 410 | 25 |

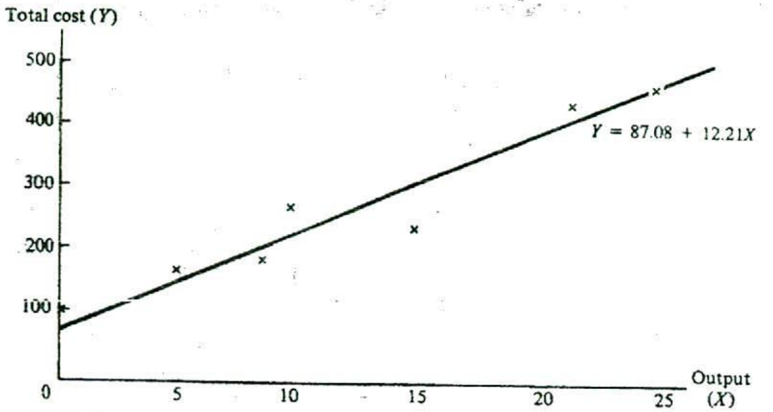



FIGURE 4.1 Total Cost, Total Output, and the Estimated Regression Equation

There is a better way. Statisticians have demonstrated that the best estimate of the coefficients of a linear function is to fit the line through the data points so that the sum of squared vertical distances from each point to the line is minimized. This technique is called *least-squares regression estimation.*

Based on the output and cost data in Table 4.1, the least-squares regression equation will be shown to be

$$\hat{Y} = 87.08 + 12.21X$$

This equation is plotted in Figure 4.1. Note that the data points fall about equally on both sides of the line.

Consider an output rate of 5. As shown in Table 4.1, the actual cost associated with this output level is 150. The value predicted by the regression equation, referred to as $\hat{Y}$, is 148.13. That is, $\hat{Y} = 87.08 + 12.21(5) = 148.13$. The deviation of the actual $Y$ value from the predicted value (i.e., the vertical distance of the point from the line), $Y_i - \hat{Y}$, is referred to as the *residual* or the *prediction error.*

There are many values that might be selected as estimators of $a$ and $b$, but only one of those sets defines a line that minimizes the sum of squared deviations [i.e., that minimizes $\Sigma(Y_i - \hat{Y}_i)^2$]. The equations for computing the least-squares estimators $\hat{a}$ and $\hat{b}$ are

$$\hat{b} = \frac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{\Sigma(X_i - \overline{X})^2} \tag{4-1}$$

and

$$\hat{a} = \overline{Y} - \hat{b}X \tag{4-2}$$

where $\overline{Y}$ and $\overline{X}$ are the means of the $Y$ and $X$ variables.

Using the basic cost and output data from the example, the necessary calculations are shown in Table 4.2. Substituting the appropriate values into equations (4-1) and (4-2), the estimates of $\hat{b}$ and $\hat{a}$ are computed to be

$$\hat{b} = \frac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{\Sigma(X_i - \overline{X})^2} = \frac{6{,}245.71}{511.40} = 12.21$$

$$\hat{a} = \overline{Y} - \hat{b}\overline{X} = 237.14 - 12.21(12.29) = 87.08$$

**TABLE 4.2 Summary Calculations for Computing the Estimates $\hat{a}$ and $\hat{b}$**

| Cost ($Y_i$) | Output ($X_i$) | $Y_i - \bar{Y}$ | $X_i - \bar{X}$ | $(X_i - \bar{X})^2$ | $(X_i - \bar{X})(Y_i - \bar{Y})$ |
|---|---|---|---|---|---|
| 100 | 0 | −137.14 | −12.29 | 151.04 | 1,685.45 |
| 150 | 5 | −87.14 | −7.29 | 53.14 | 635.25 |
| 160 | 8 | −77.14 | −4.29 | 18.40 | 330.93 |
| 240 | 10 | −002.86 | −2.29 | 5.24 | −6.55 |
| 230 | 15 | −7.14 | 2.71 | 7.34 | −19.35 |
| 370 | 23 | 132.86 | 10.71 | 114.70 | 1,422.93 |
| 410 | 25 | 172.86 | 12.71 | 161.54 | 2,197.05 |
| $\bar{Y} = 237.14$ | $\bar{X} = 12.29$ | | | $\Sigma(X_i - \bar{X})^2 = 511.40$ | $\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) = 6,245.71$ |

Thus the estimated equation for the total cost function is

$$\hat{Y} = 87.08 + 12.21X$$

The estimate of the coefficient $a$ is 87.08. This is the vertical intercept of the regression line. In the context of this example, $\hat{a} = 87.08$ is an estimate of fixed cost. Note that this estimate is subject to error because it is known that the actual fixed cost is \$100. The value of $\hat{b}$ is an estimate of the change in total cost for a one-unit change in output (i.e., marginal cost). The value of $\hat{b}$, \$12.21, means that, on average, a one-unit change in output results in a \$12.21 change in total cost. Thus $\hat{b}$ is an estimate of marginal cost.

**Key Concepts**

- The least-squares regression technique is used to estimate the coefficients of a function by fitting a line through the data so that the sum of squared deviations [i.e., $\Sigma(Y_i - \hat{Y}_i)^2$] is minimized.
- Estimates of the coefficients of the function $Y = a + bX$ are given by the equations

$$\hat{b} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \text{ and } \hat{a} = \bar{Y} - \hat{b}\bar{X}$$

- The value of $\hat{a}$ estimates the vertical intercept or the estimated value of $Y$ when $X = 0$. The value of $\hat{b}$ estimates the change in $Y$ for a one-unit change in $X$.

## Example Estimating the Demand for Lobster Dinners

The basic regression tools just discussed can also be used to estimate demand relationships. Consider a small restaurant chain specializing in fresh lobster dinners. The business has collected information on prices and the average number of meals served per day for a random sample of eight restaurants in the chain. These data are shown below. Use regression analysis to estimate the coefficients of the demand function $Qd = a + bP$. Based on the estimated equation, calculate the point price elasticity of demand at the mean values of the variables.

V.V.G

| City | Meals per Day ($Q$) | Price ($P$) |
|---|---|---|
| 1 | 100 | \$15 |
| 2 | 90 | 18 |
| 3 | 85 | 19 |
| 4 | 110 | 14 |
| 5 | 120 | 13 |
| 6 | 90 | 19 |
| 7 | 105 | 16 |
| 8 | 100 | 14 |

**Solution** The mean values of the variables are $\overline{Q} = 100$ and $\overline{P} = \$16$. The other data needed to calculate the coefficients of the demand equation are shown below.

| City | $Q_i - \overline{Q}$ | $P_i - \overline{P}$ | $(P_i - \overline{P})^2$ | $(P_i - \overline{P})(Q_i - \overline{Q})$ |
|---|---|---|---|---|
| 1 | 0 | −1 | 1 | 0 |
| 2 | −10 | 2 | 4 | −20 |
| 3 | −15 | 3 | 9 | −45 |
| 4 | 10 | −2 | 4 | −20 |
| 5 | 20 | −3 | 9 | −60 |
| 6 | −10 | 3 | 9 | −30 |
| 7 | 5 | 0 | 0 | 0 |
| 8 | 0 | −2 | 4 | 0 |
| | | | $\Sigma(P_i - \overline{P})^2 = 40$ | $\Sigma(P_i - \overline{P})(Q_i - \overline{Q}) = -175$ |

As shown, the sum of the $(P_i - \overline{P})^2$ is 40 and the sum of the $(P_i - \overline{P})(Q_i - \overline{Q})$ is −175. Thus, using equations (4-1) and (4-2),

$$\hat{b} = -175/40 = -4.375 \quad \text{and} \quad \hat{a} = 100 - (-4.375)(16) = 170.$$

Hence, the estimated demand equation is $Q_d = 170 - 4.375P$. Recall from chapter 3 that the formula for point price elasticity of demand is $E_p = (dQ/dP)(P/Q)$. Based on the estimated demand function, $dQ/dP = -4.375$. Thus, using the mean values for the price and quantity variables, $E_p = (-4.375)(16/100) = -0.7$.

## Testing Regression Estimates

Once the parameters have been estimated, the strength of the relationship between the dependent variable and the independent variables can be measured in two ways. The first uses a measure called the *coefficient of determination*, denoted as $R^2$, to measure how well the overall equation explains changes in the dependent variable. The second measure uses the $t$-statistic to test the strength of the relationship between an independent variable and the dependent variable.

**Testing Overall Explanatory Power** Define the squared deviation of any $Y_i$ from the mean of $Y$ [i.e., $(Y_i - \overline{Y})^2$] as the variation in $Y$. The total variation is found by summing these deviations for all values of the dependent variable. That is,

$$\text{total variation} = \Sigma(Y_i - \overline{Y})^2 \tag{4-3}$$

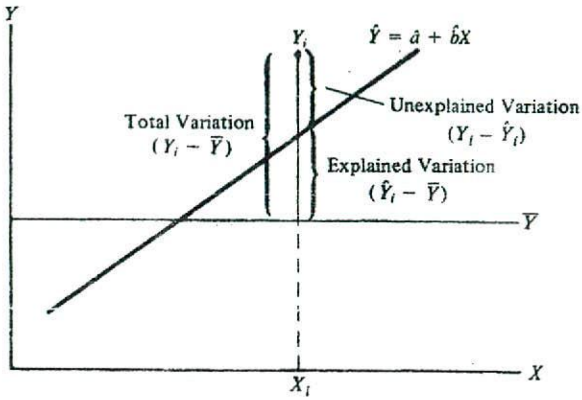



**FIGURE 4.2 Sources of Variation in a Regression Model**

Total variation can be separated into two components: explained variation and unexplained variation. These concepts are explained below. For each $X_i$ value, compute the predicted value of $Y_i$ (denoted as $\hat{Y}_i$) by substituting $X_i$ in the estimated regression equation:

$$\hat{Y}_i = \hat{a} + \hat{b}X_i$$

The squared difference between the predicted value $\hat{Y}_i$ and the mean value $\overline{Y}$ [i.e., $(\hat{Y}_i - \overline{Y})^2$] is defined as *explained variation.* The word *explained* means that the deviation of $Y$ from its average value $\overline{Y}$ is the result of (i.e., is explained by) changes in $X$. For example, in the data on total output and cost used previously, one important reason the cost values are higher or lower than $\overline{Y}$ is because output rates ($X_i$) are higher or lower than the average output rate.

Total explained variation is found by summing these squared deviations, that is,

$$\text{total explained variation} = \Sigma(\hat{Y}_i - \overline{Y})^2 \tag{4-4}$$

*Unexplained variation* is the difference between $Y_i$ and $\hat{Y}_i$. That is, part of the deviation of $Y_i$ from the average value ($\overline{Y}$) is "explained" by the independent variable, $X$. The remaining deviation, $Y_i - \hat{Y}_i$, is said to be unexplained. Summing the squares of these differences yields

$$\text{total unexplained variation} = \Sigma(Y_i - \hat{Y}_i)^2 \tag{4-5}$$

The three sources of variation are shown in Figure 4.2.

The *coefficient of determination* ($R^2$) measures the proportion of total variation in the dependent variable that is "explained" by the regression equation. That is,

$$R^2 = \frac{\text{total explained variation}}{\text{total variation}} = \frac{\Sigma(\hat{Y}_i - \overline{Y}_i)^2}{\Sigma(Y_i - \overline{Y})^2} \tag{4-6}$$

The value of $R^2$ ranges from zero to 1. If the regression equation explains none of the variation in $Y$ (i.e., there is no relationship between the independent variables and the dependent variable), $R^2$ will be zero. If the equation explains all the variation (i.e., total explained variation = total variation), the coefficient of determination will be 1. In general, the higher the value of $R^2$, the "better" the regression equation. The term *fit* is often used to describe the explanatory power of the estimated equation. When $R^2$ is high, the equation is said to fit the data well. A low $R^2$ would be indicative of a rather poor fit.

**TABLE 4.3** [illegible] Regression Model

| $Y_i$ | Total Variation $(Y_i - \overline{Y})^2$ | $\hat{Y}_i$ | Explained Variation $(\hat{Y}_i - \overline{Y})^2$ | Unexplained Variation $(Y_i - \hat{Y}_i)^2$ |
|---|---|---|---|---|
| 100 | 18,807.38 | 87.08 | 22,518.00 | 166.93 |
| 150 | 7,593.38 | 148.13 | 7,922.78 | 3.50 |
| 160 | 5,950.58 | 184.76 | 2,743.66 | 613.06 |
| 240 | 8.18 | 209.18 | 781.76 | 949.87 |
| 230 | 50.98 | 270.23 | 1,094.95 | 1,618.45 |
| 370 | 17,651.78 | 367.91 | 17,100.79 | 4.37 |
| 410 | 29,880.58 | 392.33 | 24,083.94 | 312.23 |
| $\overline{Y} = 237.14$ | $\Sigma(Y_i - \overline{Y})^2 = 79{,}942.86$ | | $\Sigma(\hat{Y}_i - \overline{Y})^2 = 76{,}245.88$ | $\Sigma(Y_i - \hat{Y}_i)^2 = 3{,}668.41$ |

How high must the coefficient of determination be in order that a regression equation be said to fit well? There is no precise answer to this question. For some relationships, such as that between consumption and income over time, one might expect $R^2$ to be at least 0.95. In other cases, such as estimating the relationship between output and average cost for fifty different producers during one production period, an $R^2$ of 0.40 or 0.50 might be regarded as quite good.

Based on the estimated regression equation for total cost and output, that is,

$$\hat{Y}_i = 87.08 + 12.21X_i$$

the coefficient of determination can be computed using the data on sources of variation shown in Table 4.3.

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{76{,}245.88}{79{,}942.86} = 0.954$$

The value of $R^2$ is 0.954, which means that more than 95 percent of the variation in total cost is explained by changes in output levels. Thus the equation would appear to fit the data quite well.

**Evaluating the Explanatory Power of Individual Independent Variables** The *t-test* is used to determine if there is a significant relationship between the dependent variable and each independent variable. This test requires that the standard deviation (or standard error) of the estimated regression coefficient be computed. The relationship between a dependent variable and an independent variable is not fixed because the estimate of $b$ will vary for different data samples. The standard error of $\hat{b}$ from one of these regression equations provides an estimate of the amount of variability in $\hat{b}$. The equation for this standard error is

$$S_{\hat{b}} = \sqrt{\frac{\Sigma(Y_i - \hat{Y}_i)^2/(n-2)}{\Sigma(X_i - \overline{X})^2}}$$

where $n$ is the number of observations. For the production-cost example used in this section, $n = 7$ and the standard error of $\hat{b}$ is

**TABLE 4.4 Selected Values of the Student's t Distribution for 95 Percent Confidence Interval**

| *Degree of Freedom* | *t-value* |
|---|---|
| 1 | 12.706 |
| 3 | 3.182 |
| 5 | 2.571 |
| 7 | 2.365 |
| 10 | 2.228 |
| 20 | 2.086 |
| 30 | 2.043 |
| 60 | 2.000 |
| 120 | 1.980 |

$$S_{\hat{b}} = \sqrt{\frac{3{,}668.41/5}{511.40}} = 1.19$$

The least-squares estimate of $\hat{b}$ is said to be an estimate of the parameter $b$. But it is known that $\hat{b}$ is subject to error and thus will differ from the true value of the parameter $b$. That is why $\hat{b}$ is called an estimate.

Because of the variability in $\hat{b}$, it sometimes is useful to determine a range or interval for the estimate of the true parameter $b$. Using principles of statistics, a 95 percent confidence interval estimate for $b$ is given by the equation

$$\hat{b} \pm t_{n-k-1} S_{\hat{b}}$$

where $t_{n-k-1}$ represents the value of a particular probability distribution known as *Student's* t *distribution.* The subscript $(n - k - 1)$ refers to the number of degrees of freedom, where $n$ is the number of observations or data points and $k$ is the number of independent variables in the equation. An abbreviated list of $t$-values for use in estimating 95 percent confidence intervals is shown in Table 4.4.[1] In the example discussed here, $n = 7$ and $k = 1$, so there are five (i.e., $7 - 1 - 1$) degrees of freedom, and the value of $t$ in the table is 2.571. Thus, in repeated estimations of the output–cost relationship, it is expected that about 95 percent of the time the true value of parameter $b$ will lie in the interval defined by the estimated value of $b$ plus or minus 2.571 times the standard error of $b$. For the output-cost data, the 95 percent confidence interval estimate would be

$$12.21 \pm 2.571(1.19)$$

or from 9.15 to 15.27. This means that the probability that the true marginal relationship between cost and output (i.e., the value of $b$) within this range is 0.95.

If there is no relationship between the dependent and an independent variable, the parameter $b$ would be zero. A standard statistical test for the strength of the relationship between $Y$ and $X$ is to check whether the 95 percent confidence interval includes

[1]A more complete list of $t$-values is found in Table III on page 668.

the value zero. If it does not, the relationship between $X$ and $Y$ as measured by $\hat{b}$ is said to be statistically *significant.* If that interval does include zero, then $\hat{b}$ is said to be *non-significant,* meaning that there does not appear to be a strong relationship between the two variables. The confidence interval for $\hat{b}$ in the output-cost example did not include zero, and thus it is said that $\hat{b}$, an estimate of marginal cost, is statistically significant or that there is a strong relationship between cost and rate of output.

Another way to make the same test is to divide the estimated coefficient ($\hat{b}$) by its standard error. The probability distribution of this ratio is the same as Student's $t$ distribution; thus this ratio is called a $t$-value. If the absolute value of this ratio is equal to or greater than the tabled value of $t$ for $n - k - 1$ degrees of freedom, $\hat{b}$ is said to be statistically significant. Using the output-cost data, the $t$-value is computed to be

$$t = \left|\frac{\hat{b}}{S_{\hat{b}}}\right| = \left|\frac{12.21}{1.19}\right| = 10.26$$

Because the ratio is greater than 2.571, the value of the $t$-statistic from Table 4.4, it is concluded that there is a statistically significant relationship between cost and output. In general, if the absolute value of the ratio $\hat{b}/S_{\hat{b}}$ is greater than the value from the table for $n - k - 1$ degrees of freedom, the coefficient $\hat{b}$ is said to be statistically significant.

**Key Concepts**

- The coefficient of determination, $R^2$, is a measure of the proportion of total variation in the dependent variable that is "explained" by the regression equation.
- A 95 percent confidence interval estimate of the parameter $b$ is given by $\hat{b} \pm t_{n-k-1}S_{\hat{b}}$. If this interval does not include zero, $\hat{b}$ is said to be statistically significant, meaning that there is a strong relationship between the dependent and independent variables.
- The ratio of an estimated regression coefficient to its standard error (the $t$-statistic) can also be used to test the statistical significance of an independent variable.

## Prediction Using Regression Equations

The regression equation can be used to predict or estimate the value of the dependent variable given the value of the independent variable.

The estimated total cost function,

$$\hat{Y} = 87.08 + 12.21X$$

can be used to make a point estimate of the cost of a particular rate of output, say 20, by substituting $X = 20$ and solving for $\hat{Y}$. Thus

$$\hat{Y} = 87.08 + 12.21(20) = 331.28$$

This means that the predicted cost of producing 20 units of output is $331.28.

Recall that the actual output-cost data points do not lie on the regression line but are dispersed above and below that line. This means that a value predicted by the regression equation will be subject to error. That is, one would not expect the predicted values to be 100 percent accurate. The standard error of the estimate ($S_e$) is a measure of the probable error in the predicted values. The formula for this standard error is

$$S_e = \sqrt{\frac{\Sigma(Y_i - \bar{Y})^2 - \hat{b}\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{n - k - 1}} \tag{4-7}$$

The predicted value $\hat{Y}$ is called a *point estimate* of the value of the dependent variable to distinguish that estimate from a confidence interval estimate. The latter is a range of values that is expected to include the actual $Y$ value 95 percent of the time. The standard error of the estimate is a fundamental part of the confidence interval estimate. For example, for a given predicted value of the dependent variable, $\hat{Y}$, the 95 percent confidence interval estimate is given by

$$\hat{Y} \pm t_{n-k-1} S_e$$

This means that 95 percent of the time the actual value of $Y$ will fall within that range.

Suppose that management is considering a production run of 22 units of output and wants to know the estimated cost. Substituting 22 for $X$ in the regression equation yields the point estimate:

$$\hat{Y} = 87.08 + 12.21(22) = 355.70$$

Using equation (4-7) and data from Tables 4.2 and 4.3, the standard error of the estimate is computed as

$$S_e = \sqrt{\frac{79{,}942.86 - (12.21)(6{,}245.71)}{5}} = 27.14$$

Thus, a 95 percent confidence interval estimate for the cost of producing 22 units of output is

$$\hat{Y} \pm 2.571(S_e) \text{ or } 355.70 \pm 2.571(27.14)$$

Based on this statistical analysis, it is expected that in repeated production runs of 22 units, 95 percent of the time the cost will be in the range \$285.92 to \$425.48. Equivalently, only 5 percent of the time will the actual cost be higher than \$425.48 or lower than \$285.92.

## *Case Study*

### The Value of a Poor Administrator

Many employees complain that their bosses do not deserve the large raises they receive each year, and college professors are no exception. If questioned, most faculty members could identify a department head, dean, vice-president, or president they believe is receiving salary adjustments that do not correspond to that person's contribution to the institution.

Are undeserved raises the exception or the rule at colleges and universities? Is there a relationship between raises received by administrators and their job performance? Faculty members at the University of South Florida used regression analysis to evaluate this question.

Each year, administrators at the university are rated by the faculty using a five-point scale, where 1 equals very poor and 5 equals very good. About 20 percent of the faculty respond to the survey. To investigate the relationship between performance and pay, ratings for 14 administrators were compared with the raises those same administrators received the same year. The equation used to estimate this relationship was $Y = B + aX$, where $Y$ is the administrator's raise and $X$ is the average rating for the individual.

Using ordinary least-squares regression, the estimated equation was $Y = 19{,}680 - 3{,}887X$. The sign of the coefficient of $X$ implies that job performance and raises were inversely related at the University of South Florida. The coefficient of $X$ was statistically significant and the $R^2$ was 0.30. The equation can be used to predict an administrator's raise based on the rating. For example, for a rating of 2, $Y = 19{,}680 - 3{,}887(2) = \$11{,}906$. The other predictions for raises are shown below.

| ***Rating*** | | ***Raise*** |
|---|---|---|
| Very poor | (1.00) | $15,793 |
| Poor | (2.00) | 11,906 |
| Average | (3.00) | 8,019 |
| Good | (4.00) | 4,132 |
| Very good | (5.00) | 245 |

One interpretation of these findings is that the university's system for awarding raises is deeply flawed. How might the low response rate to the survey affect the validity of the conclusions? ■

## Multiple Regression

Estimation of the parameters of an equation with more than one independent variable is called *multiple regression.* In principle, the concept of estimation with multiple regression is the same as with simple linear regression, but the necessary computations can be much more complicated. For an equation with three or more independent variables, the time required to calculate the values and the likelihood of an arithmetic error make computation by hand impractical. Consequently, virtually all regression analysis involving multivariate equations uses computers. There are dozens of programs available that can perform this task.

Because most economic relationships involve more than a simple relationship between a dependent and a single independent variable, multiple regression techniques are widely used in economics. For example, the demand for a product usually depends on more than just the price of the good. Other variables, such as income and prices of other goods, can also have an influence. Thus, a simple regression equation involving only quantity and price would be incomplete and probably would result in an incorrect estimation of the relationship between quantity and price. This is because the effects of other variables omitted from the equation are not taken into account. Similarly, a regression equation that included only the rate of output as the determinant of costs could generate inaccurate results because other factors, such as input prices, also affect costs. The problem of omitted variables is discussed later in this chapter.

With multiple regression, it is important that the user understand how to interpret the estimated coefficients of the equation. Earlier in the chapter it was assumed that costs were a function of output. Now assume that costs are also determined by the price of labor. Thus, the multiple regression equation can be written as

$$Y = A + bX + cZ$$

where $Y$ is total cost, $X$ is output, $Z$ is the price of labor, and $A$, $b$, and $c$ are the coefficients to be estimated. The coefficients of $X$ and $Z$ indicate the effect on total cost of a one-unit change in each variable, holding the influence of the other variable constant. For example, $b$ shows the change in total costs for a one-unit change in output, assuming that the price of labor stays the same.[2] The coefficient of $Z$ estimates the effect of a unit change in labor price, assuming that the rate of output is unchanged.

**Key Concepts**

- A regression equation can be used to predict the value of the dependent variable for given values of the independent variable.
- The standard error of the estimate is a measure of the error in prediction.
- With multiple regression, each estimated coefficient measures the impact of one variable on the dependent variable, holding constant the influence of other variables.

## DEMAND ESTIMATION

Although the process can be very complex, regression analysis really involves just four steps: (1) development of a theoretical model, (2) data collection, (3) choice of a functional form, and (4) estimation and interpretation of results. In this section, these steps are described in the context of estimating a demand equation.

### Development of a Theoretical Model

In using regression, the analyst must formulate a theoretical model of the economic relationships involved. This model should be based on sound economic theory and be expressed in mathematical terms. Fundamentally, the model-building process involves determining which variables should be included in the analysis and if there is a theoretical rationale for predicting the nature and magnitude of the relationships between variables. For example, if the objective is to estimate a demand equation, it would be reasonable to assume that quantity demanded is a function of the price of the good, income, prices of other goods, and tastes and preferences. Thus, the general form of demand equation could be written as

---

[2] In mathematical terms, $b$ is the partial derivative of $Y$ with respect to $X$. Similarly, $c$ is the partial derivative with respect to $Z$.

$$Q_d = f(P, I, P_o, T)$$

where $P$ is the price of the good, $I$ is income, $P_o$ is the price of some other good, and $T$ is a measure of tastes and preferences.

Economic theory can also be used to analyze the expected relationships between the dependent variable, $Q_d$, and the independent variables, $P$, $I$, $P_o$, and $T$. The law of demand states that price and quantity demanded are inversely related. For normal goods, income and quantity demanded would be positively related. With respect to the price of the other good, the relationship should be inverse if the two goods are complements and positive if they are substitutes. Without some additional information on the variable used to measure tastes and preferences, the nature of that association cannot be predicted.

Prior knowledge, based on economic theory, about the relationships between variables can be used to assess the empirical results of regression analysis. If the signs of the estimated coefficients are inconsistent with the predictions of economic theory, then the process should be reexamined. For example, if the estimated coefficient of price is not negative, it is possible that the demand equation was incorrectly formulated or that there were errors in data collection and entry. Other possibilities are that relevant variables were omitted from the regression equation or that there is something unique about the good or service. At the very least, a coefficient with an unexpected sign is a signal that the analysis should be carefully reexamined.

## Data Collection

To estimate demand, data for each of the variables that influence demand first must be obtained. These data may be collected from surveys, market experiments, or existing sources such as historical records of the firm or government publications. Either time-series or cross-section data may be used. Time-series data consists of period-by-period observations in a specific market for each of the variables that affect demand. Twenty months of data on quantity demanded, price, income, and prices of other goods from a single city would be an example of time-series data that could be used to estimate a demand function.

In contrast, cross-section data are based on a number of markets at a single point in time. For example, cross-section data might consist of quantity demanded, income, price, and tastes and preferences for 100 different markets during a recent year. If the cross-section data exhibit variation from market to market, they can be used to estimate relationships between quantity demanded and the other variables.

### Key Concepts

- The first step in regression analysis is to formulate a model based on economic theory.
- Economic theory can be used to evaluate the signs and magnitudes of estimated coefficients.
- Time-series data consist of observations from a single market over a period of time. Cross-section data are based on information from a number of markets at a single point in time.

## Choice of Functional Form

Equation (4-8) indicates a general relationship between quantity and the factors expected to influence demand. That is,

$$Q_d = f(P, I, P_o, T) \qquad (4\text{-}8)$$

However, estimation using regression analysis requires the choice of a specific functional form for the equation. A linear equation is the simplest possible form. The linear equation corresponding to the general relationships of equation (4-8) is

$$Q_d = B + a_p P + a_I I + a_o P_O + a_T T \qquad (4\text{-}9)$$

The linear form has several advantages. First, it can be estimated without modification—no transformations of the data are necessary. Second, the coefficients of the variables have a simple interpretation. If values of the other independent variables remain unchanged, each coefficient represents the change in quantity per unit change in the associated independent variable. Furthermore, the estimated changes are constant for each independent variable and unaffected by values of the other variables. These properties make computations much easier. For example, it is a simple matter to forecast the impact on quantity demanded of changes in per capita income. If per capita income increases by \$1,000 and $a_I = 0.05$, the increase in quantity demanded will be $0.05 \times 1{,}000$, or 50 units per period.

It is also possible to calculate elasticities based on the estimated coefficients. The formula for point elasticity of demand is

$$E_p = \frac{dQ_d}{dP} \cdot \frac{P}{Q_d}$$

But $dQ_d/dP = a_p$. Thus, by selecting a value for $P$ and using the estimated equation to compute $Q_d$, the elasticity for that price–quantity combination can be computed. Income and cross elasticities can be determined using the same general approach.

Various functional forms can be used for regression analysis. Other than the linear equation, probably the most common is the multiplicative functional form. For estimating demand, the multiplicative equation is

$$Q_d = BP^{a_p} I^{a_I} P_o^{a_o} T^{a_T} \qquad (4\text{-}10)$$

In its present form, this equation cannot be estimated using ordinary least squares because it is not linear. However, there is a simple transformation of the equation that allows it to be estimated using least squares. First, take the logarithm of both sides of equation (4-10). The result is

$$log(Q_d) = log(BP^{a_p} I^{a_I} P_o^{a_o} T^{a_T})$$

But the logarithm of a product is just the sum of the logarithms. Thus,

$$\log(Q_d) = \log(B) + \log(P^{a_p}) + \log(I^{a_I}) + \log(P_o^{a_o}) + \log(T^{a_T})$$

The equation can be further simplified by noting that the logarithm of a number raised to a power is equal to that power times the logarithm of the number. Hence,

$$\log Q_d = \log B + a_p \log P + a_I \log I + a_o log\ P_0 + a_T \log T \qquad (4\text{-}11)$$

Because this equation is linear in terms of the logarithms of the original variables, the coefficients can be estimated using the ordinary least-squares method.

The coefficients $a_p$, $a_I$, $a_o$, and $a_T$ in equations (4-10) and (4-11) have an important interpretation. Using calculus, it can easily be shown that they are the elasticities for the respective variables. Consider the price elasticity of demand. Taking the derivative of equation (4-10) with respect to price yields

$$\frac{dQ_d}{dP} = (a_p)BP^{a_p - 1}I^{a_I}P_o^{a_o}T^{a_T}$$

Multiplying both sides of the equation by $P/Q_d$ gives

$$E_p = \frac{dQ_d}{dP}\frac{P}{Q_d} = \frac{(a_p)BP^{a_p}I^{a_I}P_o^{a_o}T^{a_T}}{Q_d}$$

But

$$Q_d = BP^{a_p}I^{a_I}P_o^{a_o}T^{a_T}$$

Hence

$$E_p = a_p$$

Using an analogous approach, it can be shown that $a_I$ is the income elasticity, $a_o$ is the cross elasticity, and $a_T$ is an elasticity measure for tastes and preferences. Thus, an advantage of the multiplicative form is that it yields estimates of elasticities. Note that in contrast to the elasticities from the linear equation, these elasticities are constant—they are unaffected by changes in the independent variables.

Another feature of the multiplicative form is that the change in quantity demanded per unit change in an independent variable is not constant as it is with the linear form. Rather, it is determined not only by the associated variable, but also by the values of the other independent variables. This means computations using multiplicative equations are somewhat more difficult but they may be more realistic in depicting the relationship between the variables.

The choice of an appropriate functional form depends on the underlying theoretical model and the intended use of the results. If quantity demanded is thought to be a linear function of the independent variables, a linear form would be appropriate. In contrast, the multiplicative form may be a better choice if the objective is to estimate elasticities or to allow for nonlinear relationships between the variables.

## *Case Study*

## The Pope and the Price of Fish

For over a thousand years, the Catholic Church required that members abstain from eating meat on Fridays as an act of penance. Many Catholics responded by including fish as part of their Friday meals. The effect of this practice was to increase the demand for fish.

But in February 1966, Pope Paul VI authorized local bishops to end meatless Fridays. In December 1966, church leaders in the United States stipulated that members could eat meat on non-Lent Fridays (usually 46 Fridays during the year). With abstinence no longer required on these days, it was expected that Catholics would consume less fish and more meat. As a result, the price of fish, at least in the short run, should have decreased.

This hypothesis was examined by economist F. W. Bell. Noting that New England is 45 percent Catholic, he collected data on fish prices for seven different species consumed in that region before and after the change. He also collected data on other factors that affect fish prices, such as personal income and prices of meat and poultry. Using multiple-regression techniques to hold the effects of these other factors constant, Bell was able to assess the impact of the action by the church. His estimated equations are not reported here because they involved techniques beyond the scope of this book. However, the findings with respect to the Pope's decree are reported in the following table.

| *Species* | *Estimated Percent Decrease in Price Resulting from the Pope's Authorization to Eat Meat on Fridays* |
|---|---|
| Sea scallops | −17 |
| Yellowtail flounder | −14 |
| Large haddock | −21 |
| Small haddock (scrod) | −2 |
| Cod | −10 |
| Ocean perch | −10 |
| Whiting | −20 |

Note that the percentage decrease in prices ranged from 2 percent for small haddock to 21 percent for large haddock. The average price decline for the seven species was 12.5 percent. Clearly, many Catholics enjoyed being able to have a steak on Fridays.

SOURCE: F. W. Bell, "The Pope and the Price of Fish," *American Economic Review* (December 1968), pp. 1346–1350.

## Estimation and Interpretation of Results

It has already been noted that many software packages are available for regression analyses. Most require little more than that the analyst specify the form of the equation to be estimated and input the data. The computer then performs the necessary calculations and provides results in a format that is usually relatively easy to interpret. Consider the estimation of the demand equation for a hypothetical good. The equation to be estimated is

$$Q_d = B + a_p P + a_I I + a_o P_o$$

and the price and income variables are measured in dollars.

**TABLE 4.5 Sample Output for a Regression Problem**

| | | Variable | | |
|---|---|---|---|---|
| | *Constant* | *Price* *(P)* | *Income* *(I)* | *Price of Other Good* *($P_o$)* |
| Estimated coefficient | 50.7836 | −4.9892 | 0.0034 | −1.2801 |
| Standard error | 10.2189 | 1.3458 | 0.0045 | 0.5890 |
| $t$-statistic | (4.97) | (−3.71) | (0.76) | (−2.17) |
| | Number of observations = 182 | | $R^2 = 0.6837$ | |

Table 4.5 shows the output from the demand estimation problem. The estimated coefficient of the constant term suggests the quantity demanded if values of the other variables are zero. However, this coefficient has little economic meaning in most demand estimation problems. The other coefficients estimate the change in quantity demanded per dollar change in the associated independent variable. For example, the price coefficient is −4.9892. This estimate implies that if the other independent variables are held constant, a $1 change in the price of the good will result in a change in quantity demanded of almost five units.

In evaluating regression results, the estimated relationships between the variables should be considered in terms of economic theory. Note that the coefficient of price is negative. This implies an inverse relationship between price and quantity demanded, which is consistent with economic theory. Note also the coefficients of income and price of the other good. The signs of these coefficients could not be predicted using economic theory, but theory does provide information regarding their interpretation. The coefficient of income is positive, indicating that the item is a normal good. The "price of other good" coefficient is negative. Hence the two goods must be complements.

The standard errors reported in Table 4.5 indicate the precision of the estimates. Dividing the estimated coefficients by their standard errors gives the $t$-statistics. These can be used for hypothesis testing. A common hypothesis is that a coefficient is not significantly different from zero. As discussed earlier in the chapter, this hypothesis can be tested by comparing the computed $t$-statistics to the values shown for the $t$-distribution in Table 4.4 or in Table III on page 668.

For samples with more than about 120 observations (such as in Table 4.5), 95 percent of the $t$-distribution lies between +1.960 and −1.960. Thus $t$-statistics greater than +1.960 or less than −1.960 imply that the hypothesis that an estimated coefficient is equal to zero can be rejected with only a 5 percent probability of error. That is, the probability of erroneously concluding that a coefficient is not equal to zero is 5 percent or less.

In Table 4.5, the $t$-statistics for the constant term, price, and the price of the other good all have absolute values greater than 1.960. Thus the traditional interpretation is that these coefficients are significant (i.e., the probability of erroneously rejecting the hypothesis that they are equal to zero is less than or equal to 5 percent). In contrast, the $t$-statistic for the income variable is 0.76, meaning that there is more than a 5 percent probability of erroneously rejecting the hypothesis that the coefficient is equal to zero. Consequently, the income coefficient is referred to as nonsignificant.

Finally, the value of the coefficient of determination or $R^2$ indicates the overall explanatory power of the model. This number represents the proportion of total variation in the dependent variable explained by changes in the independent variables. The $R^2$ value from Table 4.5 is 0.6837. Thus about two-thirds of the total variation in quantity demanded is explained by price, income, and the price of the other good.

**Key Concepts**

- The functional form of an equation to be estimated should be selected based on the underlying economic theory and the intended use of the estimates.
- The estimated parameters of a linear equation indicate the impacts of a change in each independent variable.
- Without additional computation, the estimated parameters of multiplicative equations can be interpreted as elasticities.
- The estimated parameters of an equation are evaluated by examining their signs and magnitudes. The associated $t$-values are used to test hypotheses about statistical significance.

## Example Interpreting an Estimated Demand Equation

A multiplicative demand function of the form

$$Q_d = BP^{a_p}I^{a_I}P_o^{a_o}$$

is estimated using cross-section data. The results are as follows:

| | | Variable | | |
|---|---|---|---|---|
| | *Constant* | *Price* | *Income* | *Price of Other Good* |
| Estimated coefficient | 0.02248 | −0.2243 | 1.3458 | 0.1034 |
| Standard error | 0.01885 | 0.0563 | 0.5012 | 0.8145 |
| $t$-statistic | (1.19) | (−3.98) | (2.69) | (0.13) |
| | Number of observations = 224 | | $R^2 = 0.2515$ | |

1. How should the coefficients and the $R^2$ value be interpreted?
2. What will the quantity demanded be if the values of the independent variables are
   a. Price = $10.
   b. Income per capita = $9,000.
   c. Price of the other good = $15.
3. How much would quantity demanded change if price were decreased to $8 and the values of the other variables held constant?
4. Is demand elastic or inelastic? What effect would a price increase have on total revenue?
5. Are the two goods substitutes or complements?

**Solution**

1. Because the functional form is multiplicative, the estimated coefficients represent elasticities. The price elasticity is −0.2243, indicating that demand is inelastic, and the income elasticity is 1.3458, suggesting that the good is a luxury. Both of these

elasticity estimates are significantly different from zero. In contrast, the coefficient of the "other good" price is positive but not significant. The implication is that there is a weak relationship between the two goods.

The $R^2$ reported for this problem is 0.2515. This is a rather low value, even for cross-section data. It means that only about one-fourth of the variation in demand can be explained by the model. Thus forecasts generated by the estimated equation are unlikely to be very accurate. A possible explanation for the low $R^2$ is that important explanatory variables have been omitted from the model.

2. In multiplicative form, the estimated equation is

$$Q_d = 0.02248P^{-0.2243}I^{1.3458}P_o^{0.1034}$$

A forecast of demand can be obtained by substituting in the given values for the independent variables. Making these substitutions, the quantity demanded is estimated to be 3,722 units.

3. The impact of reducing price can be estimated by recomputing quantity demanded for $P = \$8$, while holding the other variables constant. The resulting $Q$ is 3,913. Hence the \$2 price decrease is estimated to increase quantity demanded by 191 units.
4. The coefficient of $P$ is the estimated price elasticity. Because $a_p = -0.2243$, demand is inelastic and a price increase would increase total revenue.
5. The coefficient of $P_o$ is positive, so the two goods are substitutes. However, the coefficient of $P_o$ is not significant.

## PROBLEMS WITH REGRESSION ANALYSIS

Although regression analysis is a valuable technique for estimating demand functions and other economic relationships, serious problems can occur if the analyst is not careful in the formulation of the model and interpretation of results. Three potential pitfalls are discussed in this section—omitted variables, the identification problem, and multicollinearity.

### Omitted Variables

It has already been observed that economic theory can be used to specify those variables that should be included in a regression equation. However, if there are variables that are omitted, the results of regression analysis can be misleading. The following example will help to explain the importance of including all the relevant variables.

It is hypothesized that the salary ($S$) of a major league baseball player depends on the number of times the player strikes out ($K$) during the season. It is assumed that the more times a player strikes out, the lower his salary would be. Hence the sign of the estimated regression coefficient is expected to be negative.

Based on data for 150 players, the following regression equation is estimated for annual salary (measured in thousands) and strikeouts in a season with the following result:

$$S = 484.42 + 15.54K \qquad R^2 = 0.44$$
$$(-5.32) \quad (2.51)$$

where the $t$-statistics are in parentheses. Note that the coefficient on strikeouts is positive and statistically significant because the $t$-statistic of 2.51 is greater than 1.96, the value of the $t$ distribution in Table 4.4, which corresponds to 148 degrees of freedom. But this result suggests the nonsensical conclusion that players are paid more if they strike out more often! Because salary is measured in thousands of dollars, the coefficient 15.54 means that each additional strikeout is associated with an additional \$15,540 in annual salary.

The problem here is that a misspecified equation has resulted in a meaningless and misleading estimate of the relationship between strikeouts and salary. A better specification would have salary dependent on strikeouts and home runs ($H$). It is hypothesized that strikeouts are negatively related with salaries and that the coefficient of home runs should be positive. Using the same data on salaries and strikeouts, but adding data on home runs, results in the following multiple-regression equation:

$$\underset{(3.71)}{S = 462.8} - \underset{(-0.33)}{1.28K} + \underset{(6.44)}{17.14H} \qquad R^2 = 0.92$$

By including information on the number of home runs, the estimated relationship between strikeouts and salary changes dramatically. That relationship now is negative (as hypothesized), although the coefficient is not significantly different from zero. However, there is a very strong relationship between salary and home runs. The estimated coefficient (17.14) suggests that each home run is associated with an additional \$17,140 in salary. Furthermore, this coefficient is statistically significant because the $t$-statistic is greater than 1.96.

Changing the specification of the equation resulted in an entirely different set of conclusions about how salary is related to performance. By including data on home runs, it is seen that the number of strikeouts is inversely related to salary. More importantly, the multiple-regression equation shows that salary is determined primarily by the number of home runs hit. Furthermore, in the first equation, $R^2$ was only 0.44, meaning that less than one-half the variation in the salary variable is explained. By including data on both strikeouts and home runs, $R^2$ increased to 0.92. The multiple-regression equation that includes both strikeouts and home runs fits the data much better.

The first equation failed to consider the number of home runs. By omitting this important information, the strange result was obtained that suggested strikeouts result in higher salaries. This erroneous conclusion resulted because many home run hitters tend to strike out quite often. Obviously, their high salaries are tied to home-run production not strikeouts. But by estimating the relationship between salary and strikeouts, it appeared that baseball players receive higher salaries for striking out more often.

The more complete analysis indicates that the number of strikeouts is relatively unimportant when the effect of home runs is considered. This suggests one of the advantages of the multiple-regression approach over simple regression. In multiple regression, the regression coefficient measures the net or partial effect of each independent variable, while holding the effect of the other independent variables constant. In the example, the coefficient on home runs is 17.14, which means that while holding the effect of strikeouts constant, each additional home run is associated with an additional \$17,140 in salary. The coefficient on strikeouts is $-1.28$. This means that after adjusting for the number of home runs, each additional strikeout is associated with \$1,280 fewer dollars of salary.

When regression results are inconsistent with economic theory, the omission of important variables could be the cause. Consider a demand function estimated from 15 quarters of time series data from a particular market. The estimating equation is $Q_d = B + aP$. Assume that the estimated coefficient for price is positive and statistically significant. One explanation for this anomalous result could be that as price increased over time, so did income and the number of people in the market. Because price is positively corre-

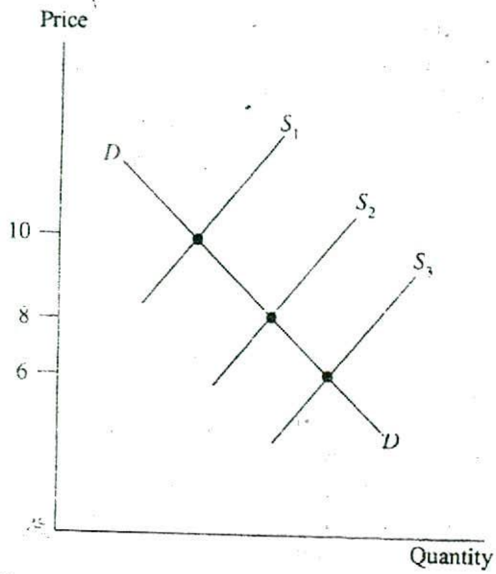



FIGURE 4.3a Single Demand Curve

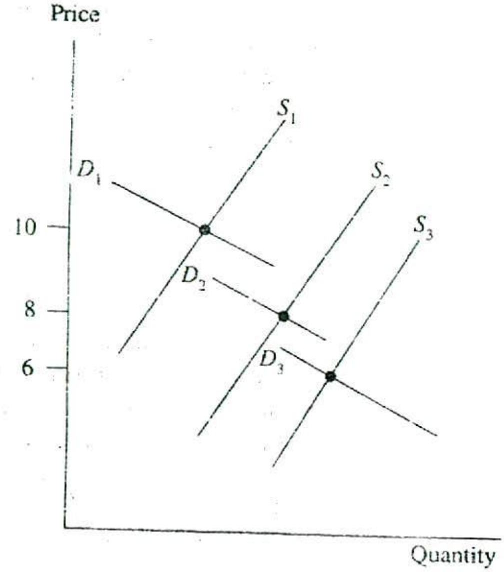



FIGURE 4.3b Multiple Demand Curves

lated with income and with population, the coefficient of the price variable is reflecting increases in demand caused by the population and income changes. To separately identify these influences, additional variables should be included in the regression equation.

## Identification

Related to the idea of omitted variables is the identification problem. Suppose that the following price and quantity information has been collected from a particular market. These data result from the interaction of supply and demand in the market. Specifically, they are the equilibrium prices and quantities observed for each year.

| *Year* | *Price* | *Quantity* |
|---|---|---|
| 1 | 10 | 100 |
| 2 | 8 | 120 |
| 3 | 6 | 140 |

In Figure 4.3a, the data have been plotted and a line fitted through the three points. Note that this line has a negative slope. If other factors that affect demand did not change over the three-year period and the supply curve shifted to the right from $S_1$ to $S_2$ to $S_3$, as shown in Figure 4.3a, then *DD* can be legitimately identified as a demand curve.

Alternatively, suppose that income, prices of other goods, or changes in tastes and preferences caused the demand curve to shift over time. In this case, the three points represent equilibrium prices and quantities determined by the intersection of different demand and supply curves, as shown in Figure 4.3b. Without more information, it is impossible to choose between the two possibilities; the separate demand curves cannot be identified.

The basic cause of this *identification problem* is that there is simultaneity between the supply and demand equations. That is, both the quantity demanded and the quantity supplied are affected by a change in the price of the good. Thus, the effects of price on quantity demanded and quantity supplied cannot be separately determined.

The key to identifying the true supply-and-demand relationships is to add one or more additional variables to each equation that are not included in the other equation. For example, assume that the problem involves the supply and demand for gasoline. The basic demand-and-supply model is

$$Q_d = B + d_1P_g$$
$$Q_s = C + s_1P_g$$
$$\text{and} \quad Q_s = Q_d$$

where $P_g$ is the price of gasoline and the equation $Q_s = Q_d$ is included because the observed data are equilibrium quantities and prices. The quantity demanded always equals the quantity supplied; thus there is no way to sort out the values of the parameters $B$, $d_1$, $C$, and $s_1$. Hence, neither function can be uniquely identified.

What is needed is another variable that affects only demand and one that affects only supply. Let income, $I$, be added to the demand equation and the price of crude oil, $P_c$, be added to the supply function. Now the three-equation model is

$$Q_d = B + d_1P_g + d_2I$$
$$Q_s = C + s_1P_g + s_2P_c$$
$$\text{and} \quad Q_s = Q_d$$

Note that the dependent variables in this model are quantity and price of gasoline (even though price appears on the right-hand side of the equations). The two independent variables are income and the price of crude oil. Together, the three equations are referred to as the *structural form* of the model.

Because $Q_d = Q_s$, the demand and supply equations can be combined to form

$$B + d_1P_g + d_2I = C + s_1P_g + s_2P_c$$

Solving this equation for $P_g$ gives

$$P_g = \frac{C - B}{d_1 - s_1} + \left\{\frac{s_2}{d_1 - s_1}\right\}P_c - \left\{\frac{d_2}{d_1 - s_1}\right\}I \tag{4-12}$$

Substituting the expression for $P_g$ back into the original demand equation and solving for quantity gives

$$Q = B + d_1\left\{\frac{C - B}{d_1 - s_1}\right\} + \left\{\frac{d_1s_2}{d_1 - s_1}\right\}P_c - \left\{\frac{d_2s_1}{d_1 - s_1}\right\}I \tag{4-13}$$

Equations (4-12) and (4-13) are called *reduced form* equations. Note that they are linear and that only the independent variables $P_c$ and $I$ appear on the righthand side. Thus, these equations can be estimated using least squares because the simultaneity problem has been eliminated. The two equations look complicated but really are not. When estimated, they each have a constant term and a coefficient for each variable. If the following substitutions are made:

$$A = B + d_1\left\{\frac{C - B}{d_1 - s_1}\right\}, \quad g_1 = \left\{\frac{d_1s_2}{d_1 - s_1}\right\}, \quad \text{and } g_2 = -\left\{\frac{d_2s_1}{d_1 - s_1}\right\}$$

then equation (4-13) can be written as

$$Q = A + g_1 P_c + g_2 I$$

The estimated parameters $A$, $g_1$, and $g_2$ combine the effects of the coefficients from the structural equations. For example, $g_1$ shows the net effect of a change in the price of crude oil on the equilibrium quantity demanded of gasoline. Equation (4-12) is interpreted in a similar manner.

When one variable on the right side of a structural equation is simultaneously determined with the variable on the left side, one approach is to determine the reduced-form equations of the model and use those equations for regression analysis. This method may be appropriate for estimating demand functions.

However, in actual practice, the simultaneity between quantity and price often is ignored, and it is assumed that the price of the good is an independent variable. In this case, an equation of the general form

$$Q_d = f(P, I, P_o)$$

is used to estimate the demand function. Any shifts in demand are captured by including the income and other goods price variables. This approach is easier than working with a simultaneous equation model, but may yield less accurate results.

## Multicollinearity

The omitted variables and the identification problem discussed thus far have both involved the need to add additional variables to the regression equation. But there are situations where the problem is that there are too many variables. An important example is when two or more independent variables are highly correlated. This problem is referred to as *multicollinearity*.

Consider a student who has just completed a difficult course in American literature. So many books were assigned in the class that no one was able to finish all of the reading. For an assignment in a statistics course, the student collected data on the grade performance and study habits of a random sample of 40 students who took the literature class. His hypothesis was that course grades should be positively related to the number of hours spent studying for the course and the amount of reading each person completed. Applying regression analysis to the data, the equation was estimated to be

$$G = \underset{(2.80)}{50.00} + \underset{(0.80)}{0.40H} + \underset{(1.35)}{0.02P} \qquad R_2 = 0.80$$

where $G$ is the percentage grade, $H$ is hours of study time, and $P$ is total number of pages read. The $t$-statistics are in parentheses. Note that the coefficients of both $H$ and $P$ are positive, but neither is statistically significant. However, the $R^2$ is very high for cross-sectional data, suggesting the equation should be a good predictor of grades.

The problem with the analysis is that hours spent studying for the literature class and pages read are likely to be highly correlated. The more time students devote to the course, the more pages they will be able to read. In fact, the relationship probably will be nearly linear. That is, twice as many hours will probably result in about twice as many pages completed.

Multicollinearity causes problems in regression analysis. When two variables are highly correlated, it is difficult to identify the unique influence that each variable has on the dependent variable. For the example, more hours spent studying mean a better grade, but more study hours also allow more pages to be read, and this also should result in a better grade.

When there is multicollinearity, the standard errors of the coefficients tend to be large and, hence, the $t$-statistics will be small. Consequently, the coefficients are less likely to be statistically significant. One solution to the problem of multicollinearity is to remove one of the highly correlated variables from the equation.

In the example, suppose that hours studied is dropped from the model. When the new equation is estimated, the results are

$$G = \underset{(2.70)}{60.00} + \underset{(3.00)}{0.03}P \qquad R^2 - 0.75$$

Note that the revised equation still has a high $R^2$ and the coefficient of $P$ is positive and statistically significant.

Multicollinearity is not always easy to detect. If two variables are almost perfectly correlated, most regression programs will indicate that they are unable to perform the analysis. But in other cases, there may be no error message, and multicollinearity may go unnoticed. One approach is to check for pairs of independent variables that are highly correlated.

**Key Concepts**

- **A misspecified regression equation can lead to erroneous conclusions about the relationship between variables.**
- **There may be an identification problem if price and quantity are simultaneously determined by supply and demand. The problem can be solved by including additional variables and using reduced-form estimation.**
- **Multicollinearity results in large standard errors and small $t$-statistics for the estimated coefficients.**

## SUMMARY

The least-squares regression technique can be used to quantify the relationship between a dependent variable and one or more independent variables. In simple regression, there is one independent variable, whereas with multiple regression, there are two or more independent variables. The estimated coefficients in a regression equation measure the change in the value of the dependent variable for each one-unit change in the independent variable, holding the other independent variables constant.

The coefficient of determination ($R^2$) is used to test the explanatory power of the entire regression equation. This statistic measures the proportion of the total variation in the dependent variable that is explained by variations in the independent variables. Hypotheses regarding the coefficients of individual independent variables are tested using the $t$-statistic, which is computed by dividing the estimated coefficient by its standard error. If the absolute value of this ratio is greater than the value taken from a table of the student's distribution, the coefficient is said to be statistically significant. Regression equations can be used to make both point and interval estimates of the predicted value of the dependent variable for given values of the independent variables.

The first step in using regression analysis is to develop a theoretical model based on economic theory. This model can specify the data that should be collected and can also be used to interpret results. Regression analysis can be applied to either cross-section or time-series data. Linear estimating equations have the advantage of simplicity, but if a multiplicative equation is used to estimate the demand function, its coefficients are elasticities.

If relevant variables are excluded from a regression equation, the equation is said to be misspecified. The use of a misspecified equation may lead to incorrect conclusions about the relationships between the dependent and independent variables. Where price and quantity are simultaneously determined by supply and demand, it may be difficult to identify the demand function. This problem can be resolved by adding explanatory variables and using a reduced-form equation. When two or more independent variables are highly correlated, this multicollinearity may result in large standard errors and, hence, small $t$-statistics for individual coefficients.

## Discussion Questions

4-1. It is sometimes argued that quantitative techniques, such as multiple regression analysis, are of little value because they are always subject to error. Is this a valid argument?

4-2. A demand equation is estimated, but the coefficient of income is not significant. In using this equation to predict quantity demanded, should the income variable be omitted? Explain.

4-3. Why would coefficients of determination tend to be higher for time-series than for cross-section data?

4-4. What is the relationship between the sign of the estimated coefficient and the sign of the associated $t$-statistic? Explain.

4-5. In a demand function, what is the economic interpretation of a statistically significant coefficient for the price variable?

4-6. In transforming a multiplicative equation to a linear form, what would be the effect on the estimated coefficient of the price variable of using natural logarithms rather than logarithms to the base 10?

4-7. Data on average price and quantity over the last 12 months are used to plot a graph that is claimed to represent a demand function. Is this a valid interpretation? Explain.

4-8. Adding more independent variables to an estimating equation tends to increase the $R^2$. What problems could result from adding variables?

4-9. Is there any basis for choosing between a linear and a multiplicative equation for estimating a demand function? Explain.

## Problems

4-1. Consider the following five data points:

| | | | | | |
|---|---|---|---|---|---|
| $X$ | −1 | 0 | 1 | 2 | 3 |
| $Y$ | −1 | 1 | 2 | 4 | 5 |

a. Use regression analysis to calculate by hand the estimated coefficients of the equation $Y = B + aX$.
b. Compute the standard error and the $t$-statistic for the coefficient of $X$.

4-2. Consider the following five data points:

| | | | | | |
|---|---|---|---|---|---|
| $X$ | −1.0 | 0.0 | 1.0 | 2.0 | 3.0 |
| $Y$ | −1.0 | 1.0 | 1.0 | 2.5 | 3.5 |

a. Use regression analysis to calculate by hand the estimated coefficients of the equation $Y = B + aX$.
b. Compute the coefficient of determination.
c. What is the predicted value of $Y$ for $X = 1.0$? For $X = 3.5$?

4-3. The following regression equation estimates the relationship between the number of cups of hot chocolate sold ($H$) and number of swimmers ($N$) at a beach:

$$H = 252.8 - 2.05N \qquad R^2 = 0.45$$
$$(2.06) \quad (-3.05)$$

($t$-values are shown in parentheses.)

a. Explain or interpret the regression coefficient of $N$, the $t$-value, and the coefficient of determination of this equation.
b. How is it possible that more hot chocolate is sold when there are fewer people at the beach? Does this relationship suggest anything about the specification of the equation?

4-4. The vice-president of United Feeds, Inc., has provided the following quarterly price–quantity data for UF Superb, a horse feed additive. He has asked that the demand equation for this product be estimated.

| | *Quarter* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
| Price ($) | 60 | 53 | 43 | 40 | 47 | 57 | 41 | 53 | 37 | 51 |
| Quantity | 83 | 93 | 100 | 108 | 97 | 80 | 105 | 86 | 110 | 90 |

a. Use the ordinary least-squares regression method to estimate this function.
b. The firm is considering a price increase to $66. Make a point and 95 percent confidence interval estimate of sales volume if this price increase is carried out.

b 90% T
Point

4-5. Annual prices and beef consumption per capita in six cities are as follows:

| City | *Price per Pound* | *Consumption per Capita* |
|---|---|---|
| 1 | $2.00 | 55 |
| 2 | 1.90 | 60 |
| 3 | 2.10 | 50 |
| 4 | 1.80 | 70 |
| 5 | 2.30 | 45 |
| 6 | 2.20 | 48 |

a. If a demand equation is to be estimated using these data, would the linear form (i.e., $Q = B + aP$) or the multiplicative form (i.e., $Q = BP^a$) be more appropriate? Explain.

b. Could the resulting equation be properly interpreted as a demand function? Explain.

4-6. The MacWend Drive-In has determined that demand for hamburgers is given by the following equation:

$$\underset{(1.85)}{Q = 205.2} + \underset{(2.64)}{23.0A} - \underset{(-5.61)}{200.0P_M} + \underset{(2.02)}{100.0P_C} + \underset{(4.25)}{0.5I}$$

where $Q$ is the number of hamburgers sold per month (in 1,000s), $A$ is the advertising expenditures during the previous month (in \$1,000s), $P_M$ is the price of MacWend burgers (dollars), $P_C$ is the price of hamburgers of the company's major competitor (dollars), and $I$ is income per capita in the surrounding community (in \$1,000s). The $t$-statistic for each coefficient is shown in parentheses below each coefficient.

a. Are the signs of the individual coefficients consistent with predictions from economic theory? Explain.

b. If $A$ = \$5,000, $P_M$ = \$1, $P_c$ = \$1.20, and $I$ = \$20,000, how many hamburgers will be demanded?

c. What is the advertising elasticity at $A$ = \$5,000?

4-7. Motorland Recreational Vehicles estimates the monthly demand ($Q$) for its product is given by the equation

$$\log Q = \underset{(1.20)}{1.00} - \underset{(-2.50)}{1.50 \log P} + \underset{(0.02)}{3.00 \log I} \qquad R^2 = 0.21$$

where $P$ is price and $I$ is income per capita in thousands. The $t$-statistics are shown in parentheses and logarithms to the base 10 were used to transform the equation. Assume that estimates are generated by a sample of 400 observations.

a. Rewrite the expression as a multiplicative demand equation.

b. Based on the equation, is the product an inferior good, a necessity, or a luxury good? How much confidence do you have in your answer? Explain.

c. Is the equation likely to be useful in predicting demand for Motorland's product? Why or why not?

4-8. Data from 20 cities were used to estimate the demand for face-lifts. The resulting regression equation was

$$Q_d = \underset{(5.42)}{50.000} - \underset{(-2.34)}{0.001P} + \underset{(2.00)}{0.002I} \qquad R^2 = 0.55$$

where $Q_d$ is face-lifts per 1,000 population per year, $P$ is the price in dollars, $I$ is income in dollars, and the $t$-statistics are shown in parentheses.

a. In determining statistical significance of the coefficients, what number should be used for degrees of freedom?

b. Which of the coefficients are statistically significant at the 5 percent level?

c. The mean values of $P$ and $I$ are \$5,000 and \$20,000, respectively. Compute the point price elasticity of demand.

d. What is the predicted demand at the mean values of the independent variables?

4-9. A demand equation of the form $Q = BP^aI^b$ is estimated, and the coefficient of price is $-2$ with a standard error of 0.8. The coefficient of income is 3 with a standard error of 2.0. The estimate is based on a sample size of 43.

a. How many degrees of freedom should be used in determining the critical value of the $t$-statistic? What is the table value for the 95 percent confidence level?

b. Which of the coefficients are statistically significant at the 95 percent confidence level? Explain.

4-10. For a hypothesis test of the statistical significance of coefficients, what would be the degrees of freedom for each of the following equations? What is the table value of the $t$-statistic for the 95 percent confidence level?

a. $Q = BP^aI^bP^c$ and 11 observations?

b. $Y = A + bX + cZ$ and 63 observations?

c. $Y = A + bX + cZ$ and 200 observations?

## Problems Requiring Calculus

4.11. The demand function for a product is estimated to be

$$Q_d = 10P^{-1.0}I^{2.0}P_o^{0.0}$$

and the mean values for $P$ and $I$ are \$150 and \$18,000, respectively.

a. Predict the value for the dependent variables at the mean values of the independent variables.

b. At the mean values, how much will a dollar change in price change the quantity demanded? What about a dollar change in income?

4-12. Economists on a skiing vacation in Aspen, Colorado, decide to study the demand for lift tickets in the area. Using multiple regression analysis, the demand equation is estimated to be

$$\log Q_t = 3.00 - 2.0 \log P_t + 0.5 \log S$$

where $Q_t$ is quantity demanded per pay, $P_t$ is the price of a full-day lift ticket, and $S$ is accumulated snow depth in inches at the lodges. Using calculus, show that for small changes, a 1 percent increase in accumulated snow depth will increase daily lift ticket demand by 0.5 percent.

4-13. Starting with the general demand equation,

$$Q = bP_p^{a}I_I^{a}P_o^{a}$$

demonstrate that $a_I$ is the income elasticity and $a_O$ is the cross elasticity.

## Computer Problems

The following problems can be solved by using the TOOLS program (downloadable from www.prenhall.com/petersen) or by using other computer software.

4-14. Data on income (in thousands of dollars). education (years), experience (years), and age (years) for twenty people are shown here.

| *Person* | *Income* | *Age* | *Education* | *Job Experience* |
|---|---|---|---|---|
| 1 | 5.0 | 29 | 2 | 9 |
| 2 | 9.7 | 36 | 4 | 18 |
| 3 | 28.4 | 41 | 8 | 21 |
| 4 | 8.8 | 30 | 8 | 12 |
| 5 | 21.0 | 34 | 8 | 14 |
| 6 | 26.6 | 36 | 10 | 16 |
| 7 | 25.4 | 61 | 12 | 16 |
| 8 | 23.1 | 29 | 12 | 9 |
| 9 | 22.5 | 54 | 12 | 18 |
| 10 | 19.5 | 30 | 12 | 5 |
| 11 | 21.7 | 28 | 12 | 7 |
| 12 | 24.8 | 29 | 13 | 9 |
| 13 | 30.1 | 35 | 14 | 12 |
| 14 | 24.8 | 59 | 14 | 17 |
| 15 | 28.5 | 65 | 15 | 19 |
| 16 | 26.0 | 30 | 15 | 6 |
| 17 | 38.9 | 40 | 16 | 17 |
| 18 | 22.1 | 23 | 16 | 1 |
| 19 | 33.1 | 58 | 17 | 10 |
| 20 | 48.3 | 60 | 21 | 17 |

a. Use multiple regression analysis to estimate income as a linear function of age. Write the equation, *t*-statistics, and the coefficient of determination. Provide an explanation for the sign of the age coefficient.
b. Use regression analysis to estimate income as a linear function of education, job experience, and age. Write the equation, *t*-statistics, and the coefficient of determination. How do the results of part (b) explain the results from part (a)?
c. Use the results from part (b) to estimate income of a typical person who has 14 years of education, has 10 years of job experience, and is 45 years of age.

4-15. Data on grade point average and IQ were obtained for 12 high school students.

| *Grade Point Average* | *IQ* | *Grade Point Average* | *IQ* |
|---|---|---|---|
| 2.1 | 116 | 2.9 | 126 |
| 2.2 | 129 | 2.7 | 122 |
| 3.1 | 123 | 2.1 | 114 |
| 2.3 | 121 | 1.7 | 109 |
| 3.4 | 131 | 3.3 | 132 |
| 2.9 | 134 | 3.5 | 140 |

a. Use regression analysis to estimate the effect of IQ on grade point average. Write the equation, *t*-statistics, and the coefficient of determination. Is the result consistent with your prior expectations? Explain.
b. Forecast the grade point average for a student with an IQ of 120 and for a student with an IQ of 150. Which forecast do you have more confidence in? Why?

4-16. Data on electric power consumption (in billions of kilowatt-hours), GNP (in billions of dollars), and electricity prices (in cents per kilowatt-hour) for the period 1969–1983 are shown here.

| *Year* | *Consumption* | *GNP* | *Price* |
|---|---|---|---|
| 1969 | 407.9 | $ 944.0 | 2.09¢ |
| 1970 | 447.8 | 992.7 | 2.10 |
| 1971 | 479.1 | 1,077.6 | 2.19 |
| 1972 | 511.4 | 1,185.9 | 2.29 |
| 1973 | 554.2 | 1,326.4 | 2.38 |
| 1974 | 555.0 | 1,434.2 | 2.83 |
| 1975 | 586.1 | 1,594.2 | 3.21 |
| 1976 | 613.1 | 1,718.0 | 3.45 |
| 1977 | 652.3 | 1,918.3 | 3.78 |
| 1978 | 679.2 | 2,163.9 | 4.03 |
| 1979 | 696.0 | 2,417.8 | 4.43 |
| 1980 | 734.4 | 2,631.7 | 5.12 |
| 1981 | 730.5 | 2,957.8 | 5.80 |
| 1982 | 732.7 | 3,069.3 | 6.44 |
| 1983 | 750.9 | 3,304.8 | 6.83 |

a. Using regression analysis, estimate consumption as a linear function of GNP, price, and the previous year's electricity consumption. (NOTE: assume 1968 consumption was $367.7 billion.) Write the equation, *t*-statistics, and the coefficient of determination. Are the signs of the estimated coefficients consistent with economic theory? Which of the coefficients are statistically significant at the 0.05 level?

b. In 1984, GNP was $3661.3 billion and the price of electricity was 7.16 cents per kilowatt-hour. Use the estimating equation from part (a) to predict electricity consumption for 1984.

4-17. Consumption of hamburgers (thousands of burgers per week) in 12 different cities is shown here. Prices of hamburgers, income per capita (in $1,000s), and prices of hot dogs for the cities are also shown.

| *City* | *Hamburger Consumption* | *Hamburger Price* | *Income ($1,000s)* | *Hot Dog Price* |
|---|---|---|---|---|
| 1 | 50 | $1.50 | 12.0 | $1.80 |
| 2 | 80 | 1.35 | 14.2 | 1.55 |
| 3 | 95 | 1.25 | 15.0 | 1.45 |
| 4 | 105 | 1.20 | 16.0 | 1.35 |
| 5 | 70 | 1.40 | 13.8 | 1.60 |
| 6 | 85 | 1.30 | 14.3 | 1.50 |
| 7 | 55 | 1.50 | 13.3 | 1.70 |
| 8 | 60 | 1.45 | 13.3 | 1.70 |
| 9 | 75 | 1.35 | 13.7 | 1.60 |
| 10 | 90 | 1.25 | 14.5 | 1.50 |
| 11 | 100 | 1.20 | 15.2 | 1.35 |
| 12 | 65 | 1.45 | 13.6 | 1.65 |

a. Use regression analysis to estimate hamburger consumption as a multiplicative function of the price of hamburgers, income, and hot dog price. Write the equation, $t$-statistics, and the coefficient of determination. Which coefficients are significant at the 0.05 level?

b. Based on the estimates from part (a), what are the price, income, and cross elasticities? Is the cross elasticity consistent with economic theory? Explain.

CHAPTER

# 5

# Business and Economic Forecasting

## PREVIEW

The vast majority of business decisions involve some degree of uncertainty—managers seldom know exactly what the outcomes of their choices will be. One approach to reducing the uncertainty associated with decision making is to devote resources to forecasting. Forecasting involves predicting future economic conditions and assessing their effect on the operations of the firm.

Frequently, the objective of forecasting is to predict demand. In some cases, managers are interested in the total demand for a product. For example, the decision by an office products firm to enter the home computer market may be determined by estimates of industry sales growth. In other circumstances, the projection may focus on the firm's probable market share. If a forecast suggests that sales growth by existing firms will make successful entry unlikely, the company may decide to look for other areas in which to expand.

Forecasts can also provide information on the proper product mix. For an automobile manufacturer such as General Motors, managers must determine the number of full-sized versus compact cars to be produced. In the short run, this decision is largely constrained by the firm's existing production facilities for producing each kind of car. However, over a longer period, managers can build or modify production facilities. But such choices must be made long before the vehicles begin coming off the assembly line. Accurate forecasts can reduce the uncertainty caused by this long lead time. For example, if the price of gasoline is expected to increase, the relative demand for compact cars is also likely to increase. Conversely, a projection of stable or falling gasoline prices might stimulate demand for larger cars.

Forecasting is an important management activity. Major decisions in large businesses are almost always based on forecasts of some type. In some cases, the forecast may be little more than an intuitive assessment of the future by those involved in the decision. In other circumstances, the forecast may have required thousands of work hours and tens of thousands of dollars. It may have been generated by the firm's own economists, provided by consultants specializing in forecasting, or be based on information provided by government agencies.

This chapter focuses on some basic techniques of forecasting. The first section considers various methods for collecting the necessary data. Next is a discussion of time-series analysis. The third section considers barometric forecasting. The basic principles of input/output analysis are presented in the last section of the chapter.

## SOURCES OF DATA

Forecasting requires the development of a good set of data on which to base the analysis. A forecast cannot be better than the data from which it is derived. Three important sources of data used in forecasting are expert opinion, surveys, and market experiments.

### Expert Opinion

The collective judgment of knowledgeable persons can be an important source of information. In fact, some forecasts are made almost entirely on the basis of the personal insights of key decision makers. This process may involve managers conferring to develop projections based on their assessment of the economic conditions facing the firm. In other circumstances, the company's sales personnel may be asked to evaluate future prospects. In still other cases, consultants may be employed to develop forecasts based on their knowledge of the industry. Although predictions by experts are not always the product of "hard data," their usefulness should not be underestimated. Indeed, the insights of those closely connected with an industry can be of great value in forecasting.

Methods exist for enhancing the value of information elicited from experts. One of the most useful is the *Delphi technique.* Its use can be illustrated by a simple example. Suppose that a panel of six outside experts is asked to forecast a firm's sales for the next year. Working independently, two panel members forecast an 8 percent increase, three members predict a 5 percent increase, and one person predicts no increase in sales. Based on the responses of the other individuals, each expert is then asked to make a revised sales forecast. Some of those expecting rapid sales growth may, based on the judgments of their peers, present less optimistic forecasts in the second iteration. Conversely, some of those predicting slow growth may adjust their responses upward. However, there may also be some panel members who decide that no adjustment of their initial forecast is warranted.

Assume that a second set of predictions by the panel includes one estimate of a 2 percent sales increase, one of 5 percent, two of 6 percent, and two of 7 percent. The experts again are shown each other's responses and asked to consider their forecasts further. This process continues until a consensus is reached or until further iterations generate little or no change in sales estimates.

The value of the Delphi technique is that it aids individual panel members in assessing their forecasts. Implicitly, they are forced to consider why their judgment differs from that of other experts. Ideally, this evaluation process should generate more precise forecasts with each iteration.

One problem with the Delphi method can be its expense. The usefulness of expert opinion depends on the skill and insight of the experts employed to make predictions. Frequently, the most knowledgeable people in an industry are in a position to command large fees for their work as consultants. Or they may be employed by the firm, but have other important responsibilities, which means that there can be a significant opportunity cost in involving them in the planning process. Another potential problem is that those who consider themselves experts may be unwilling to be influenced by the predictions of others on the panel. As a result, there may be few changes in subsequent rounds of forecasts.

### Surveys

Surveys of managerial plans can be an important source of data for forecasting. The rationale for conducting such surveys is that plans generally form the basis for future actions. For example, capital expenditure budgets for large corporations are usually planned well in advance. Thus, a survey of investment plans by such corporations should provide a reasonably accurate forecast of future demand for capital goods.

Several private and government organizations conduct periodic surveys of plant and equipment expenditure plans. One of the most widely used is sponsored by McGraw-Hill, Inc. This survey is conducted twice yearly and includes corporations accounting for more than 50 percent of total investment in the U.S. economy. An even more comprehensive survey of capital expenditure plans is undertaken quarterly by the U.S. Department of Commerce. The results of this survey are reported in the department's *Survey of Current Business.*

Useful data for forecasting can also be obtained from surveys of consumer plans. For example, the Survey Research Center at the University of Michigan polls consumers about their intentions to purchase specific products such as household appliances, housing, and automobiles. The results are used to project consumer demand and also to measure the level of consumer confidence in the economy. The U.S. Bureau of the Census also conducts surveys of consumer intentions.

If data from existing sources do not meet its specific needs, a firm may conduct its own survey. Perhaps the most common example involves companies that are considering a new product or making a substantial change in an existing product. But with new or modified products, there are no data on which to base a forecast. One possibility is to survey households regarding their anticipated demand for the product. Typically, such surveys attempt to ascertain the demographic characteristics (e.g., age, education, and income) of those who are most likely to buy the product and find how their decisions would be affected by different pricing policies.

Although surveys of consumer demand can provide useful data for forecasting, their value is highly dependent on the skills of their originators. Meaningful surveys require careful attention to each phase of the process. Questions must be precisely worded to avoid ambiguity. The survey sample must be properly selected so that responses will be representative of all customers. Finally, the methods of survey administration should produce a high response rate and avoid biasing the answers of those surveyed. Poorly phrased questions or a nonrandom sample may result in data that are of little value.

Even the most carefully designed surveys do not always predict consumer demand with great accuracy. In some cases, respondents do not have enough information to determine if they would purchase a product. In other situations, those surveyed may be pressed for time and be unwilling to devote much thought to their answers. Sometimes the response may reflect a desire (either conscious or unconscious) to put oneself in a favorable light or to gain approval from those conducting the survey. Because of these limitations, forecasts seldom rely entirely on results of consumer surveys. Rather, these data are considered supplemental sources of information for decision making.

## *Case Study*
### The Use and Abuse of Survey Data

For years, firms have used consumer surveys to collect data on the demand for their products. A more recent trend has been the use of survey data to promote those products to potential customers. But where advocacy and persuasion are the goals, there may be a tendency to stretch the limits of good survey design and interpretation.

In some cases, the problem with a survey is that the questions have the effect of biasing the answers in a way that favors the products of the firm. A study sponsored by

Levi Strauss asked college students to select which clothes they thought would be most popular during the coming year. The company announced that Levi's 501s were chosen as the most popular jeans. What the public was not told was that 501s were the only jeans on the list.

A Black Flag survey stated that, "A roach disk . . . poisons a roach slowly. The dying roach returns to the nest and after it dies is eaten by other roaches. In turn, these roaches become poisoned and die." Survey respondents were then asked, "How effective do you think this type of product would be in killing roaches?" Provided with this "helpful" information, 79 percent said that the disk would be effective.

A sample that is not random can be used to generate survey results that are suspect. A Chrysler study showed that its cars were preferred to those of Toyota. However, none of the people in the sample owned a foreign car, suggesting that they may have been predisposed to buy U.S. automobiles. A survey sponsored by American Express and the French government concluded that it was untrue that the French are unfriendly. But the sample consisted of Americans who had visited France for pleasure more than once during the last two years—presumably people who have positive feelings about vacationing in France.

Sample bias was the cause of what is generally considered to be the most inaccurate political poll in U.S. history. In 1936, the *Literary Digest* predicted that Republican Alf Landon would be a big winner over Franklin D. Roosevelt in the presidential election of that year. But the sample consisted of those who had telephones, cars, or were subscribers to the magazine—all characteristics of high-income voters at that time. History records that Roosevelt was reelected to a second term by a landslide vote of the general population.

The selective use of data is not confined to Western society. In China, a population census determined that the population of one province was 28 million. Five years later, the same province was found to have 105 million people. An astounding birth rate was not the cause. Rather, the first census was used for military conscription (hence, a small number was better) and the second was the basis for government aid to the victims of a famine in the regions (now a large estimated population was beneficial). ■

## Market Experiments

A potential problem with survey data is that survey responses may not translate into actual consumer behavior. That is, consumers do not necessarily do what they say they are going to do. This weakness can be partially overcome by the use of market experiments designed to generate data prior to the full-scale introduction of a product or implementation of a policy.

To set up a market experiment, the firm first selects a test market. This market may consist of several cities, a region of the country, or a sample of consumers taken from a mailing list. Once the market has been selected, the experiment may incorporate a number of features. It may involve evaluating consumer perceptions of a new product in the test market. In other cases, different prices for an existing product might be set in various cities in order to determine demand elasticity. A third possibility would be a test of consumer reaction to a new advertising campaign.

There are several factors that managers should consider in selecting a test market. First, the location should be of manageable size. If the area is too large, it may be expensive and difficult to conduct the experiment and to analyze the data. Second, the residents of the test market should resemble the overall population of the United States in age, education, and income. If not, the results may not be applicable to other areas. Finally, it should be possible to purchase advertising that is directed only to those who are being tested.

Market experiments have an advantage over surveys in that they reflect actual consumer behavior, but they still have limitations. One problem is the risk involved. In test markets where prices are increased, consumers may switch to products of competitors. Once the experiment has ended and the price reduced to its original level, it may be difficult to regain those customers. Another problem is that the firm cannot control all the factors that affect demand. The results of some market experiments can be influenced by bad weather, changing economic conditions, or the tactics of competitors. Finally, because most experiments are of relatively short duration, consumers may not be completely aware of pricing or advertising changes. Thus their responses may understate the probable impact of those changes.

**Key Concepts**

- The Delphi technique is an iterative technique that can be used to enhance the value of expert opinion.
- Surveys of consumer intentions require careful attention to wording, sample selection, and methods of administration.
- Market experiments provide actual data on consumer behavior. However, they are expensive, risky, and may be influenced by factors beyond the firm's control.

## TIME-SERIES ANALYSIS

Regression analysis, as described in chapter 4, can be used to quantify relationships between variables. However, data collection can be a problem if the regression model includes a large number of independent variables. When changes in a variable show discernable patterns over time, time-series analysis is an alternative method for forecasting future values.

The focus of time-series analysis is to identify the components of change in the data. Traditionally, these components are divided into four categories:

1. Trend
2. Seasonality
3. Cyclical patterns
4. Random fluctuations

A *trend* is a long-term increase or decrease in the variable. For example, the time series of human population in the United States exhibits an upward trend, while the trend for endangered species, such as the tiger, is downward. The *seasonal* component represents changes that occur at regular intervals. A large increase in sales of skis in the fall and early winter would be an example of seasonality.

Analysis of a time series may suggest that there are *cyclical patterns*, defined as sustained periods of high values followed by low values. Business cycles fit this category. Finally, the remaining variation in a variable that does not follow any discernable pattern is due to *random fluctuations*. Various methods can be used to determine trends, seasonality, and any cyclical patterns in time-series data. However, by definition, changes in the variable due to random factors are not predictable. The larger the random component of a time series, the less accurate the forecasts based on those data.

## Trend Projection

One of the most commonly used forecasting techniques is trend projection. As the name suggests, this approach is based on the assumption that there is an identifiable trend in a time series of data. Trend projection can also be used as the starting point for identifying seasonal and cyclical variations.

Table 5.1 is a time series of a firm's quarterly sales over a three-year time span. These data will be used to illustrate graphical and statistical trend projection and also to describe a method for making seasonal adjustments to a forecast.

**Graphic Curve Fitting** Note that the data show generally increasing sales quarter by quarter. But a useful forecast usually requires greater precision than is implied by the statement "generally increasing sales." To be of value in forecasting, a numerical estimate of the increase in sales per quarter must be made. One way to make this estimate is to fit a line to the data graphically. In Figure 5.1, a straight line is drawn through the data points in such a way as to reflect the trend of the data as accurately as possible.

The upward slope of the line reflects sales increases over time. By computing the slope of this trend line, it is possible to determine the average rate of increase per quarter. This value can be used to calculate sales in future periods. Alternatively, by extending the trend line beyond the last data point (i.e., the fourth quarter of 1998), the esti-

**TABLE 5.1 Hypothetical Time-Series Sales Data**

| *Period Number* | *Quarter* | *Sales (millions)* |
|---|---|---|
| 1 | 1996:I | $300 |
| 2 | 1996:II | 305 |
| 3 | 1996:III | 315 |
| 4 | 1996:IV | 340 |
| 5 | 1997:I | 346 |
| 6 | 1997:II | 352 |
| 7 | 1997:III | 364 |
| 8 | 1997:IV | 390 |
| 9 | 1998:I | 397 |
| 10 | 1998:II | 404 |
| 11 | 1998:III | 418 |
| 12 | 1998:IV | 445 |

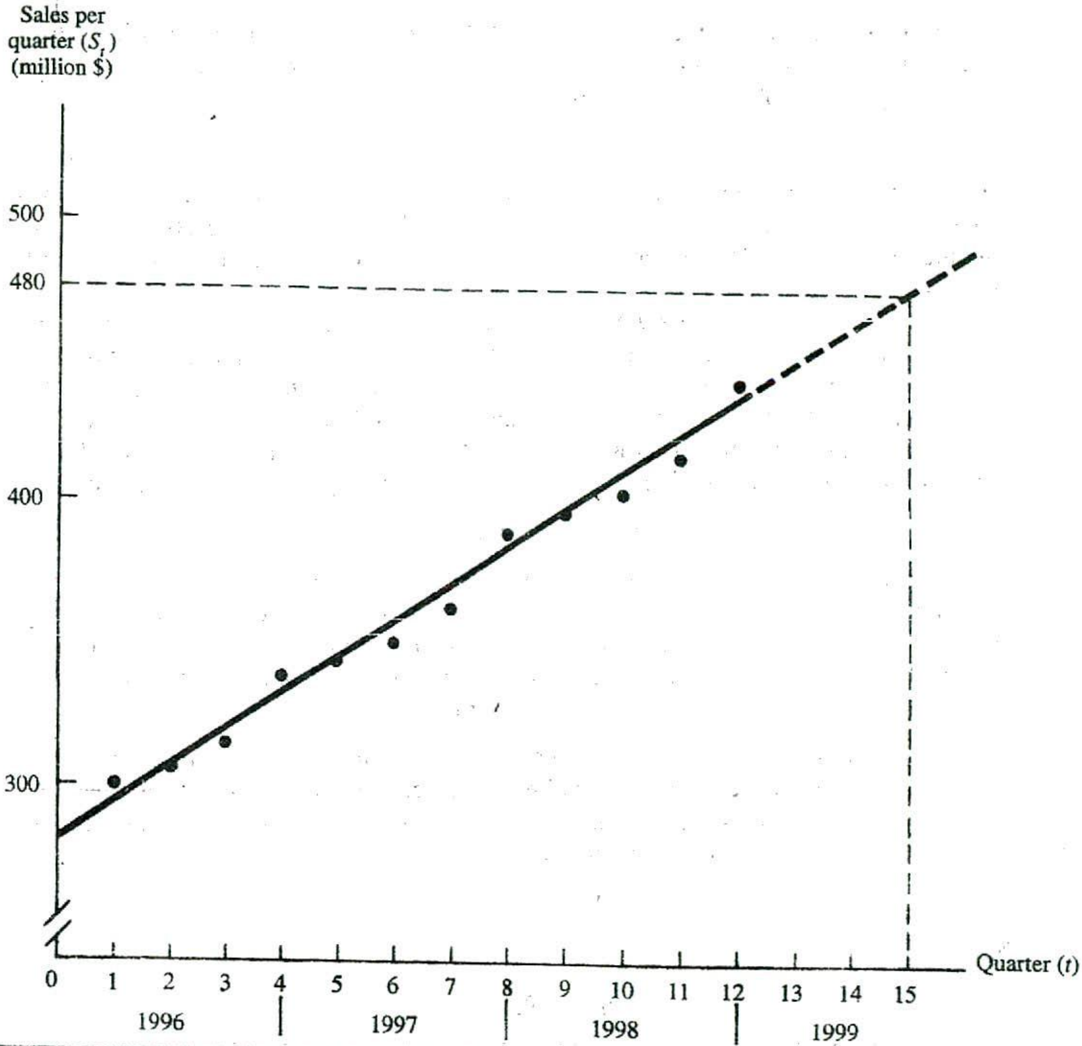



FIGURE 5.1 Curve Fitting of Time-Series Data

mated sales can be read directly from the graph. Suppose that sales are to be forecast for the third quarter of 1999. Based on an extrapolation of the data to that period, the graph indicates that sales are projected to be about $480 million.

**Statistical Curve Fitting** One limitation of graphical curve fitting should be obvious—the accuracy of forecast depends on the analyst's ability to fit a curve to the data. A more sophisticated approach is to use statistical methods to fit the data to an equation of specified functional form. Basically, this involves using the ordinary least-squares concept developed in chapter 4 to estimate the parameters of the equation.

CONSTANT RATE OF CHANGE Suppose that an analyst determines that a forecast will be made assuming that there will be a constant rate of change in sales from one period to the next. That is, the firm's sales will change by the same amount between two periods. The time-series data of Table 5.1 are to be used to estimate that rate of change. Statistically, this involves estimating the parameters of the equation

$$S_t = S_0 + bt \tag{5-1}$$

where $S$ denotes sales and $t$ indicates the time period. The two parameters to be estimated are $S_0$ and $b$. The value of $S_0$ corresponds to the sales (vertical) axis intercept of the line in Figure 5.1. The parameter $b$ is the constant rate of change and corresponds to the slope of the line in Figure 5.1.

Many hand calculators can estimate the parameters of equation (5-1). Specific procedures vary from model to model, but usually the only requirement is that the user input the data and push one or two designated keys. The machine then returns the estimated parameters. For the data of Table 5.1, the quarters would have to be inputted as sequential numbers starting with 1. That is, 1996: I would be entered as 1, 1996: II would be entered as 2, and so forth. Based on the data from the table, equation (5-1) is estimated as

$$S_t = 281.394 + 12.811t$$

The interpretation of the equation is that the estimated constant rate of increase in sales per quarter is \$12.811 million. A forecast of sales for any future quarter, $S_t$, can be obtained by substituting in the appropriate value for $t$. For example, the third quarter of 1999 is the 15th observation of the time series. Thus, the estimated sales for that quarter would be 281.394 + 12.811(15), or \$473.56 million.

CONSTANT PERCENTAGE RATE OF CHANGE Now suppose that the individual responsible for the forecast wants to estimate a percentage rate of change in sales. That is, it is assumed that sales will increase by a constant percent each period. This relationship can be expressed mathematically as

$$S_t = S_{t-1}(1 + g)$$

Similarly,

$$S_{t-1} = S_{t-2}(1 + g)$$

where $g$ is the constant percentage rate of change, or the growth rate. These two equations imply that

$$S_t = S_{t-2}(1 + g)^2$$

and, in general,

$$S_t = S_0(1 + g)^t \tag{5-2}$$

As shown, the parameters of equation (5-2) cannot be estimated using ordinary least squares. The problem is that the equation is not linear. However, there is a simple transformation of the equation that allows it to be estimated using ordinary least squares.

First, take the natural logarithm of equation (5-2).[1] The result is

[1]Either natural logarithms or logarithms to the base 10 can be used. In this section, natural logarithms (designated as ln) are used for all computations.

$$\ln S_t = \ln[S_0(1 + g)^t]$$

But the logarithm of a product is just the sum of the logarithms. Thus

$$\ln S_t = \ln S_0 + \ln[(1 + g)^t]$$

The right-hand side of the equation can be further simplified by noting that

$$\ln[(1 + g)^t] = t[\ln(1 + g)]$$

Hence

$$\ln S_t = \ln S_0 + t[\ln(1 + g)] \tag{5-3}$$

Equation (5-3) is linear in form. This can be seen by making the following substitutions:

$$Y_t = \ln S_t$$
$$Y_0 = \ln S_0$$
$$b = \ln(1 + g)$$

Thus the new equation is

$$Y_t = Y_0 + bt \tag{5-4}$$

which is linear.

The parameters of equation (5-4) can easily be estimated using a hand calculator. The key is to recognize that the sales data have been translated into logarithms. Thus, instead of $S_t$, it is $\ln S_t$ that must be entered as data. However, note that the $t$ values have not been transformed. Hence for the first quarter of 1996, the data to be entered are $\ln 300 = 5.704$ and 1; for the second quarter, $\ln 305 = 5.720$ and 2; and so forth. The transformed data are provided in Table 5.2.

**TABLE 5.2 Natural Logarithms of Hypothetical Time-Series Sales Data**

| *Period Number* ($t$) | *Quarter* | *Natural Logarithm of Sales (in millions)* ($S_t$) |
|---|---|---|
| 1 | 1996:I | 5.704 |
| 2 | 1996:II | 5.720 |
| 3 | 1996:III | 5.753 |
| 4 | 1996:IV | 5.829 |
| 5 | 1997:I | 5.847 |
| 6 | 1997:II | 5.864 |
| 7 | 1997:III | 5.897 |
| 8 | 1997:IV | 5.966 |
| 9 | 1998:I | 5.984 |
| 10 | 1998:II | 6.001 |
| 11 | 1998:III | 6.036 |
| 12 | 1998:IV | 6.098 |

Using the ordinary least-squares method, the estimated parameters of equation (5-4), based on the data from Table 5.2 are

$$Y_t = 5.6623 + 0.0353t$$

But these parameters are generated from the logarithms of the data. Thus, for interpretation in terms of the original data, they must be converted based on the relationships $\ln S_0 = Y_0 = 5.6623$ and $\ln(1 + g) = b = 0.0353$. Taking the antilogs yields $S_0 = 287.810$ and $1 + g = 1.0359$. Substituting these values for $S_0$ and $1 + g$ back into equation (5-2) gives

$$S_t = 287.810(1.0359)^t$$

where 287.810 is sales (in millions of dollars) in period 0 and the estimated growth rate, $g$, is 0.0359 or 3.59 percent.

To forecast sales in a future quarter, the appropriate value of $t$ is substituted into the equation. For example, predicted sales in the third quarter of 1999 (i.e., the fifteenth quarter) would be $287.810(1.0359)^{15}$, or \$488.51 million.

**Seasonal Variation in Time-Series Data** Seasonal fluctuations in time-series data are not uncommon. In particular, a large increase in sales for the fourth quarter is a characteristic of certain industries. Indeed, some retailing firms make nearly half of their total sales during the Christmas holiday period. Other business activities have their own seasonal sales patterns. Electric utilities serving hot, humid areas have distinct peak sales periods during the summer months because of the extensive use of air conditioning, whereas those in colder regions may have peaks in winter. Similarly, housing sales drop off during the winter, but the demand for accountants' services increases in the first quarter as income tax deadlines approach.

A close examination of the data in Table 5.1 on page 150 indicates that the quarterly sales increases are not uniformly distributed over the year. The increases from the first quarter to the second, and from the fourth quarter to the first, tend to be small, while the fourth-quarter increase is consistently larger than that of other quarters. That is, the data exhibit seasonal fluctuations, as shown by the dashed line in Figure 5.2.

Pronounced seasonal variation can cause serious errors in forecasts based on time-series data. For example, Table 5.1 indicates that actual sales for the fourth quarter of 1998 were \$445 million. But if the estimated equation is used to predict sales for that period (using the constant rate of change model), the predicted total is 281.394 + 12.811(12), or \$435.13 million. The large difference between actual and predicted sales occurs because the equation does not take into account the fourth-quarter sales jump. Rather, the predicted value from the equation represents an averaging of individual quarters. Thus, sales will be underestimated for the strong fourth quarter. Conversely, the predicting equation may overestimate sales for the other quarters.

The accuracy of the forecast can be improved by seasonally adjusting the data. Probably the most common method of adjustment is the *ratio-to-trend* approach. Its use can be illustrated using the data from Table 5.1. Based on the predicting equation,

$$S_t = 281.394 + 12.811t$$

actual and calculated fourth-quarter sales are shown in Table 5.3 on page 156. The final column of the table is the ratio of actual to predicted sales for the fourth quarter. This ratio is a measure of the seasonal error in the forecast.

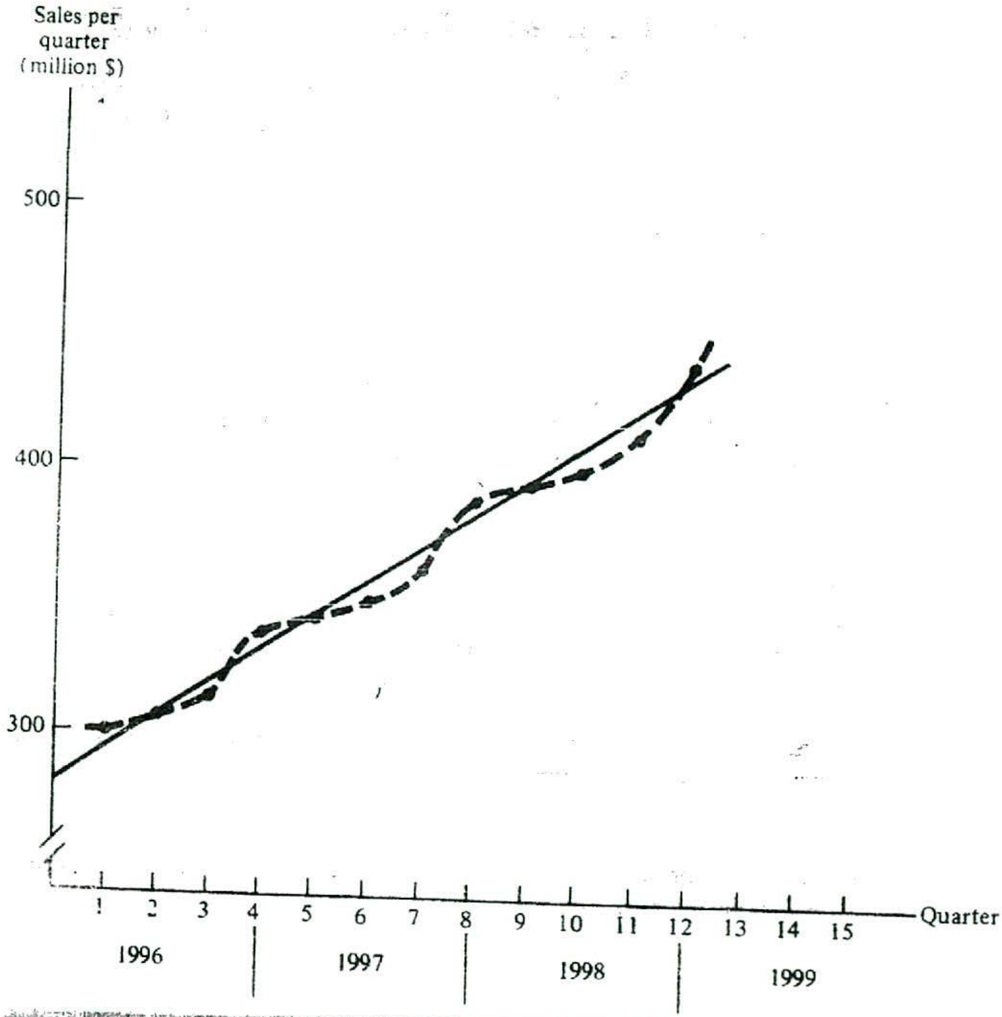



FIGURE 5.2 Seasonal Variation in Time-Series Data

As shown, for the three-year period, average actual sales for the fourth quarter were 102 percent of the average forecasted sales for that quarter. The factor 1.02 can be used to adjust future fourth-quarter sales estimates. For example, if the objective is to predict sales for the fourth quarter of 1998, the predicting equation generates an estimate of $435.13 million. Multiplying this number by the 1.020 adjustment factor, the forecast is increased to $443.8 million, which is close to the actual sales of $445 million for that quarter. A similar technique could be used to make a downward adjustment for predicted sales in other quarters.

Seasonal adjustment can improve forecasts based on trend projection. However, trend projection still has some shortcomings. One is that it is primarily limited to short-term predictions. If the trend is extrapolated much beyond the last data point, the accuracy of

**TABLE 6.5 Seasonal Adjustment Using the Ratio-to-Trend Method**

| Year | *Forecasted Fourth-Quarter Sales* | *Actual Fourth-Quarter Sales* | *Actual/Predicted Fourth-Quarter Sales* |
|---|---|---|---|
| 1996 | 332.64 | $340 | 1.022 |
| 1997 | 383.88 | 390 | 1.016 |
| 1998 | 435.13 | 445 | 1.023 |
| | | | Average = 1.020 |

the forecast diminishes rapidly. Another limitation is that factors such as changes in relative prices and fluctuations in the rate of economic growth are not considered. Rather, the trend projection approach assumes that historical relationships will not change.

**Key Concepts**

- Forecasts based on the assumption of a constant rate of change can be made by fitting time-series data to an equation of the general form

$$S_t = S_0 + bt$$

- Forecasts assuming a constant percentage rate of change involve estimating the parameters of the equation

$$S_t = S_0(1 + g)^t$$

- Forecast errors due to seasonal variation can be reduced using the ratio-to-trend adjustment method.

# *Case Study*

## Forecasting Winning Performances in the Olympic Games

The Summer Olympics brings together the world's best athletes. In most events, there has been a general trend over the years for winning times, distances, weights, and scores to improve as training methods and equipment are perfected and more athletes become involved in the sport. The following table indicates the winning distance for the women's shot put and the winning time for the men's 400-meter race for the 12 Olympics held between 1948 and 1992.

| *Year* | *Winning Distance or Time* | |
|---|---|---|
| | *Women's Shot Put (meters)* | *Men's 400 Meters (seconds)* |
| 1948 | 13.75 | 46.2 |
| 1952 | 15.28 | 45.9 |
| 1956 | 16.59 | 46.7 |
| 1960 | 17.32 | 44.9 |
| 1964 | 18.14 | 45.1 |
| 1968 | 19.61 | 43.8 |
| 1972 | 21.03 | 44.7 |
| 1976 | 21.16 | 44.3 |
| 1980 | 22.40 | 44.6 |
| 1984 | 20.48 | 44.3 |
| 1988 | 22.24 | 43.9 |
| 1992 | 21.06 | 43.5 |

Trend projection can be used to estimate the rate of improvement and forecast gold-medal performances for future Olympic games. If the constant rate and percentage rate of change models are used to analyze the data, the results are as follows:

Women's Shot put

$Y_t = 14.403 + 0.721t$ — Forecast for 1996: 23.77 meters

$Y_t = 14.585(1.041)^t$ — Forecast for 19 24.44 meters

Men's 400 Meters

$Y_t = 46.359 - 0.236t$ — Forecast for 1996: 43.29 seconds

$Y_t = 46.369(0.995)^t$ — Forecast for 1996: 43.31 seconds

The analysis implies that the winning distance for the women's shot put has increased about seven-tenths of a meter, or 4 percent per Olympics. The gold medal 400-meter time has decreased by about two-tenths of a second, or approximately one-half of one percent during each four-year interval.

The actual winning times for the 1996 Olympics in Atlanta, Georgia, were 20.55 meters for the shot put and 43.5 seconds for the 400 meters. The trend projection equations for the 400 meters predict the 1996 winning time within two-tenths of a second, but the forecasts for the women's shotput are much too high. The problem is that the rapid increases in distance in the earlier years (e.g., 1948 to 1972) have not been matched in more recent Olympics. In fact, the winning shotput distance for 1996 was about the same as in 1972. One possible explanation is that the gender and drug tests that are now required of athletes have reduced cheating in the field events. The failure of the shot put equations to accurately predict 1996 results indicates the need for the analyst to be aware of any conditions or events that could influence time-series data. ■

## Exponential Smoothing

Trend projection is actually just regression analysis where the only independent variable is time. One characteristic of this method is that each observation has the same weight. That is, the effect of the initial data point on the estimated coefficients is just as great as the last data point. If there has been little or no change in the pattern over the entire time series, this is not a problem. However, in some cases, more recent observations will contain more accurate information about the future than those at the beginning of the series. For example, the sales history of the last three months may be more relevant in forecasting future sales than data for sales 10 years in the past.

*Exponential smoothing* is a technique of time-series forecasting that gives greater weight to more recent observations. The first step is to choose a smoothing constant, $\alpha$, where $0 < \alpha < 1.0$. If there are $n$ observations in a time series, the forecast for the next period (i.e., $n + 1$) is calculated as a weighted average of the observed value of the series at period $n$ and the forecasted value for that same period. That is,

$$F_{n+1} = \alpha X_n + (12 - \alpha)F_n \tag{5-5}$$

where $F_{n+1}$ is the forecast value for the next period, $X_n$ is the observed value for the last observation, and $F_n$ is a forecast of the value for the last period in the time series. The forecasted values for $F_n$ and all the earlier periods are calculated in the same manner. Specifically,

$$F_t = \alpha X_{t-1} + (1 - \alpha)F_{t-1} \tag{5-6}$$

starting with the second observation (i.e., $t = 2$) and going to the last (i.e., $t = n$). Note that equation (5-6) cannot be used to forecast $F_1$ because there is no $X_0$ or $F_0$. This problem is usually solved by assuming that the forecast for the first period is equal to the observed value for that period. That is, $F_1 = X_1$. Using equation (5-6), it can be seen that this implies that the second-period forecast is just the observed value for the first period, or $F_2 = X_1$.

The exponential smoothing constant chosen determines the weight that is given to different observations in the time series. As $\alpha$ approaches 1.0, more recent observations are given greater weight. For example, if $\alpha = 1.0$, then $(1 - \alpha) = 0$ and equations (5-5) and (5-6) indicate that the forecast is determined only by the actual observation for the last period. In contrast, lower values for $\alpha$ give greater weight to observations from previous periods.

Assume that a firm's sales over the last 10 weeks are as shown in Table 5.4. By assumption, $F_2 = F_1 = X_1$. If $\alpha = 0.20$, then

$$F_3 = 0.20(4.30) + 0.80(400) = 406.0$$

and

$$F_4 = 0.20(420) + 0.80(406) = 408.8$$

The forecasted values for four different values of $\alpha$ are provided in Table 5.4. The table also shows forecasted sales for the next period after the end of the time-series data, or week 11. Using $\alpha = 0.20$, the forecasted sales value for the 11th week is computed to be

$$F_{11} = 0.20(420) + 0.80(435.7) = 432.56$$

**TABLE 5.4 Forecasts Based on Exponential Smoothing**

| *Week* $t$ | *Sales* $X_t$ | $\alpha = 0.20$ $F_t$ | $\alpha = 0.40$ $F_t$ | $\alpha = 0.60$ $F_t$ | $\alpha = 0.80$ $F_t$ |
|---|---|---|---|---|---|
| 1 | 400 | 400.00 | 400.00 | 400.00 | 400.00 |
| 2 | 430 | 400.00 | 400.00 | 400.00 | 400.00 |
| 3 | 420 | 406.00 | 412.00 | 418.00 | 424.00 |
| 4 | 440 | 408.80 | 415.20 | 419.20 | 420.80 |
| 5 | 460 | 415.04 | 425.12 | 431.68 | 436.18 |
| 6 | 440 | 424.03 | 439.07 | 448.67 | 455.23 |
| 7 | 470 | 427.23 | 439.44 | 443.47 | 443.05 |
| 8 | 430 | 435.78 | 451.67 | 459.39 | 464.61 |
| 9 | 440 | 434.62 | 443.00 | 441.76 | 436.92 |
| 10 | 420 | 435.70 | 441.80 | 440.70 | 439.38 |
| 11 | — | 432.56 | 433.08 | 428.28 | 423.88 |

Table 5.4 suggests why this method is referred to as smoothing technique. Consider the forecasts based on $\alpha = 0.20$. Note that the smoothed data show much less fluctuation than the original sales data. Note also that as $\alpha$ increases, the fluctuations in the $F_t$ increase, because the forecasts give more weight to the last observed value in the time series.

**Choice of a Smoothing Constant** Any value of $\alpha$ could be used as the smoothing constant. One criterion for selecting this value might be the analyst's intuitive judgment regarding the weight that should be given to more recent data points. But there is also an empirical basis for selecting the value of $\alpha$. Remember that the coefficients of a regression equation are chosen to minimize the sum of squared deviations between observed and predicted values. This same method can be used to determine the smoothing constant.

The term $(X_t - F_t)^2$ is the square of the deviation between the actual time-series data and the forecast for the same period. Thus, by adding these values for each observation, the sum of the squared deviations can be computed as

$$\sum_{t=1}^{n} (X_t - F_t)^2$$

One approach to choosing $\alpha$ is to select the value that minimizes this sum. For the data and values of $\alpha$ shown in Table 5.4, these sums are

| *Smoothing Constant* | *Sum of Squared Deviations* |
|---|---|
| 0.20 | 6484.23 |
| 0.40 | 4683.87 |
| 0.60 | 4213.08 |
| 0.80 | 4394.52 |

These results suggest that, of the four values of the smoothing constant, $\alpha = 0.60$ provides the best forecasts using these data. However, it should be noted that there may be values of $\alpha$ between 0.60 and 0.80 or between 0.40 and 0.60 that yield even better results.

**Evaluation of Exponential Smoothing** One advantage of exponential smoothing is that it allows more recent data to be given greater weight in analyzing time-series data. Another is that, as additional observations become available, it is easy to update the forecasts. There is no need to reestimate the equations, as would be required with trend projection.

The primary disadvantage of exponential smoothing is that it does not provide very accurate forecasts if there is a significant trend in the data. If the time trend is positive, forecasts based on exponential smoothing will be likely to be too low, while a negative time trend will result in estimates that are too high. Simple exponential smoothing works best when there is no discernable time trend in the data. There are, however, more sophisticated forms of exponential smoothing that allow both trends and seasonality to be accounted for in making forecasts.[2]

# *Case Study*

## Forecasting Economic Growth and the Federal Deficit

Among the hundreds of time series collected and published by the federal government, two of the most important are the rate of real economic growth and the federal deficit. The numbers for these two series during the 1980s are shown here.

| *Year* | *Growth Rate in Real GDP (%)* | *Deficit (billions)* |
|---|---|---|
| 1980 | −0.2 | 61.3 |
| 1981 | 1.9 | 63.8 |
| 1982 | −2.5 | 145.9 |
| 1983 | 3.6 | 176.0 |
| 1984 | 6.8 | 169.6 |
| 1985 | 3.4 | 196.6 |
| 1986 | 2.7 | 206.9 |
| 1987 | 3.4 | 158.2 |
| 1988 | 4.5 | 141.7 |
| 1989 | 2.5 | 134.3 |

Based on these data, the growth rates and deficits for future years can be forecast using either trend projection or exponential smoothing. The estimated equations for the constant rate-of-change trend projection model are

$$\text{Growth}_t = 0.413 + 0.399t$$
$$\text{Deficit}_t = 93.913 + 8.385t$$

Using exponential smoothing, a smoothing constant is selected by examining the sum of squared deviations between the observed and forecasted values. For the 1980s

[2]See, for example, S. Makridakis and S. Wheelwright, *Forecasting Methods for Management* (New York: Wiley, 1989), pp. 71–94

data, it was found that the sum of squared deviations is minimized using $\alpha = 0.4$ for the growth rate and $\alpha = 1.0$ for the deficit. The choice of $\alpha = 1.0$ for the deficit data implies that, with exponential smoothing, the best forecast of the next period is just the observed value for the previous period. In contrast, using growth data, the forecast for the last period is weighted more heavily than the actual value for the previous period.

The actual and forecasted values for 1990 are provided below. Note that none of the forecasts are close to the observed values for 1990. One explanation is that growth rates and deficits don't follow a regular pattern over time. Rather, they are determined by a variety of factors, such as fiscal and monetary policy and world economic conditions. As such, regression models that incorporate other variables should generate more accurate forecasts. However, even using very sophisticated econometric models, the track record for forecasting growth rates and deficits is not very impressive. ■

| | *Growth Rate* | *Deficit* |
|---|---|---|
| Trend projection forecast | 4.8 | 191.6 |
| Exponential smoothing forecast | 2.9 | 134.3 |
| Actual value for 1990 | 0.9 | 161.3 |

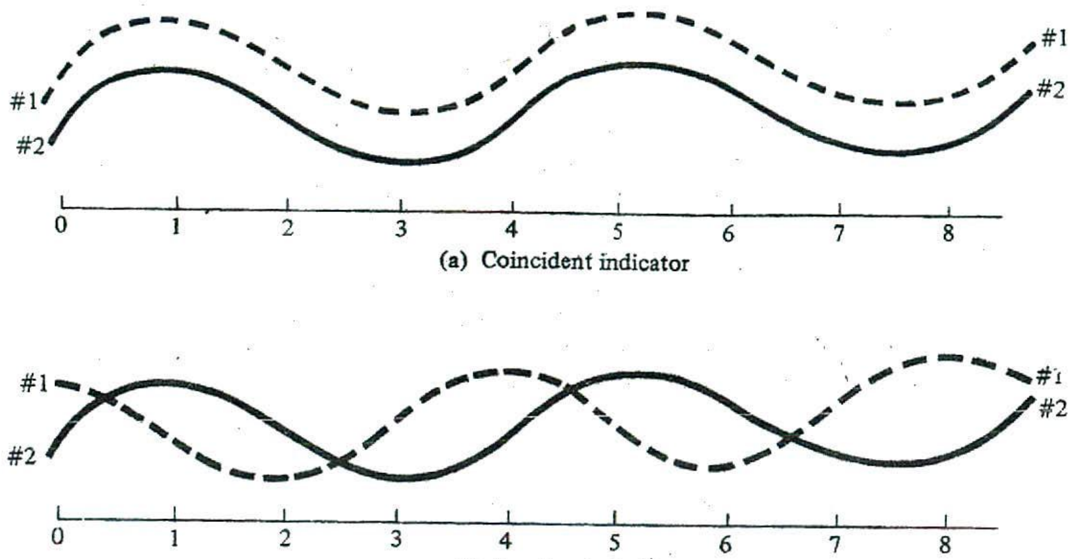
**Key Concepts**

- Forecasts based on exponential smoothing are a weighted average of observed and predicted values for the previous period.
- Weights are determined by selecting a smoothing constant that minimizes the sum of squared deviations between the forecasted and observed values.

## BAROMETRIC FORECASTING

Trend projection and exponential smoothing use time-series data to predict the future based on past relationships. But if there is no clear pattern in a time series, the data are of little value for forecasting. An alternative approach is to find a second series of data that is correlated with the first. Hence, by observing changes in the second series, it may be possible to predict changes in the first. For example, suppose that a lumber company's sales exhibited large yearly fluctuations over the last decade—so large that any forecast based on a trend projection of sales is useless. But over the same period, it is also determined that the firm's sales were highly correlated with the number of housing starts. Thus, if housing starts can be predicted, this information can be used to forecast lumber sales.

A time series that is correlated with another time series is sometimes called an *indicator* of the second series. Substantial time and effort have been expended searching




(a) Coincident indicator

(b) Leading indicator

for good indicators of economic trends. *Business Conditions Digest,* a monthly publication of the Department of Commerce, reports information on over 300 time series.[3] These data are closely followed by economists, managers, and financial analysts.

## Leading Indicators

If two series of data frequently increase or decrease at the same time, one series may be regarded as a coincident indicator of the other. For example, in Figure 5.3a, series 1 is a coincident indicator of series 2 because the two series have their peaks and troughs in the same periods.

If changes in one series consistently occur prior to changes in another series, a leading indicator has been identified. In Figure 5.3b, series 1 can be considered a leading indicator of series 2 because the peaks and troughs of series 1 consistently occur before the corresponding peaks and troughs of series 2.

For purposes of forecasting, leading indicators are of primary interest. Much as a meteorologist uses changes in barometric pressure to predict the weather, leading indicators can be used to forecast changes in general economic conditions. Consequently, the use of such indicators is commonly referred to as *barometric forecasting.*

The value of a leading indicator depends on several factors. First, the indicator must be accurate. That is, its fluctuations must correlate closely with fluctuations in the series that it is intended to predict. Second, the indicator must provide adequate lead time. Even if two series are highly correlated, an indicator will be of little use if the lead time is too short. For example, because of the time needed to put idle steel furnaces into operation,

---

[3] U.S. Department of Commerce, *Business Conditions Digest* (Washington, D.C.: U.S. Government Printing Office, monthly).

a series that predicts changes in steel demand by, say, one week would not be a useful forecasting tool for managers in the steel industry. A third requirement is that the lead time be relatively constant. If a series leads another by six months on one occasion and by two years the next, the indicator will be of little use because it cannot provide useful forecasts.

Fourth, there should be a logical explanation as to why one series predicts another. If enough time series are studied, it is likely that a correlation can be found between pairs of series. However, unless there is a causal relationship between the two series, the historical pattern may not be very useful in forecasting future events because there is no reason to expect the pattern to be repeated. For example, it has been suggested that skirt lengths are a leading indicator of stock prices. Specifically, shorter skirts predict higher prices as measured by the Dow Jones Industrial Average. Some attempts have been made to explain this relationship, but they are not very convincing. Finally, an indicator's value is affected by the cost and time necessary for data collection. A time series that can be maintained only at a very high cost may not be worth the expense. Similarly, if there is a long delay before the data are available, the effective lead time of the indicator may be too short to be useful.

Table 5.5 lists selected leading indicators and an economic variable that each is used to predict. For some of the indicators, there is an obvious link between the two series. One example is new building permits and private housing starts. When a permit is issued, this indicates a strong commitment to build a new house. Thus changes in permits should be closely correlated to changes in housing starts. Clearly, housing permits represent a leading indicator because builders are legally required to have a permit before beginning construction. Similar arguments can be made to explain new orders as leading indicators of sales in the durable goods and capital equipment industries.

For some of the indicators, it is more difficult to explain the correlation between the two series. Consider stock prices as a predictor of general economic conditions. Historically, indices of common stock prices have been a relatively accurate predictor of cycles in business activity. One possible explanation is that stock prices reflect expectations and plans of managers and consumers, which are implemented in future months.

## Composite and Diffusion Indices

Although a time series showing the changes in stock prices may be somewhat useful in predicting general economic conditions, no single leading indicator has yet been identified

**TABLE 5.5 Selected Leading Indicators** [illegible]

| *Leading Indicator* | *Economic Variable Predicted by the Indicator* |
|---|---|
| 1. Average workweek | Manufacturing output |
| 2. Average weekly initial unemployment claims | State unemployment insurance payments |
| 3. New orders for durable goods | Sales of durable goods |
| 4. New orders for capital goods | Sales of capital goods |
| 5. New building permits | Private housing starts |
| 6. Change in manufacturing and trade inventories | General economic conditions |
| 7. Industrial material prices | Consumer prices |
| 8. Common stock prices | General economic conditions |

that comes close to having a perfect forecasting record. In addition, even when an indicator correctly predicts variations in economic activity, the lead time often shows considerable fluctuation. The basic problem is that any time series contains random fluctuations that do not conform to the general pattern of the data.

One way to improve barometric forecasting is by construction of an index. Such indices represent a single time series made up of a number of individual leading indicators. The purpose of combining the data is to smooth out the random fluctuations in each individual series. Ideally, the resulting index should provide more accurate forecasts. The most commonly used barometric forecasting indices are composite indices and diffusion indices.

**Composite Indices** A composite index is a weighted average of individual indicators. The weights are based on the predictive ability of each series. That is, a series that does a better job of predicting would be given greater weight than a less accurate series. The index is interpreted in terms of percentage changes from period to period. For example, the most widely followed composite index is maintained by the U.S. Department of Commerce and is based on 11 leading indicators.

Each month the department reports the change in the index. Monthly swings have little significance, but sustained increases for several months are interpreted as a sign that general economic conditions are likely to improve. Conversely, successive declines in the index suggest that the economy is weakening.

**Diffusion Indices** The same 11 leading indicators also are used by the Department of Commerce to construct a diffusion index. This index is a measure of the proportion of the individual time series that increase from one month to the next. For example, if eight of the indicators increased from June to July, the diffusion index for July would be 8/11, or 72.7 percent. When the index is over 50 percent for several months, the forecast is for improved economic conditions. As the index approaches 100 percent, the likelihood of improvement increases. Conversely, if less than 50 percent of the indicators exhibit an increase, a downturn is indicated.

The use of indices improves the accuracy of barometric forecasting. However, the prediction record of this technique is far from perfect. On several occasions the Department of Commerce indices have forecast recessions that failed to occur. Variability in lead time is another weakness. A third problem is that, while the barometric approach signals the likely direction of changes in economic conditions, it says little about the magnitude of such changes. As such, it provides only a qualitative forecast. Finally, although there has been much study on the leading indicators of general economic conditions, the managers of individual firms may find it difficult to identify leading indicators that provide accurate forecasts for their specific needs.

**Key Concepts**

- The value of a leading indicator depends on the accuracy of prediction, length and stability of the lead time, and the cost and availability of data.
- Composite indices consist of a weighted average of several leading indicators.
- A diffusion index is calculated as the proportion of indicators that increase from one period to the next.

## Example Calculating Composite and Diffusion Indices

Following are data for three leading indicators for a three-month period. The first month represents the base period, and the three series are to be given equal weight. Construct a composite and a diffusion index from the data.

| *Month* | *Leading Indicator I* | *Leading Indicator II* | *Leading Indicator III* |
|---|---|---|---|
| 1 | 400 | 30 | 100 |
| 2 | 425 | 29 | 110 |
| 3 | 460 | 33 | 135 |

**Solution** The diffusion index is generated by determining whether each series increased or decreased from month to month. For the second month, series I and III increased, but series II decreased. Hence the index is 66.7 for that month. During the third month, all of the series increased in comparison to month 2. Thus the index for that month is 100.

The composite index can be computed by first calculating the percentage changes (relative to the base month) for each series. Percentage changes during the second month were 6.25 percent for the first series, −3.33 percent for the second, and 10 percent for the third. Giving each series equal weight, the average percentage change was 4.31 percent. The value of the composite index for the first month is arbitrarily set to 100, so the index for the second month is 104.31. For the third month, the changes from the base period are 60/400 = 15 percent, 3/30 = 10 percent, and 35/100 = 35 percent, respectively. Thus, the average change is 20.0 percent. Hence the index for that month is 120.0.

Values for the two indices for each month are shown in the following table. Both indices suggest that economic conditions should improve in future months.

| *Month* | *Diffusion Index* | *Composite Index* |
|---|---|---|
| 1 | — | 100.00 |
| 2 | 66.70 | 104.31 |
| 3 | 100.00 | 120.00 |

# INPUT/OUTPUT ANALYSIS

Econometric models can be used to forecast changes in demand in one sector, but such models seldom have the detail necessary to assess the impacts of those changes on other sectors. However, modern economic systems are highly interrelated. Changes in demand in one sector of the economy can have significant impacts on demand in other sectors. Some of these impacts are immediate and obvious. For example, steel, rubber, glass, and plastics are all important inputs in the production of motor vehicles. Thus, an increase in the demand for automobiles would cause an increase in the demand for those four products.

The direct impact that automobile sales have on the demand for steel, rubber, glass, and plastics is augmented by secondary effects generated in still other sectors. Consider the resulting increase in demand for steel. As steel output increases to meet the requirements for producing automobiles, steel industry managers will find it necessary to purchase additional inputs. These may include iron ore, coal, and electricity. If the increase in demand is expected to be permanent, management may also decide to expand the capacity of their production facility by acquiring additional capital goods such as blast furnaces.

In turn, these secondary impacts will affect other parts of the economy. Over time, the increase in automobile demand may be the cause of change in hundreds of different industries. To the casual observer, many of these changes would seem unrelated to the production of automobiles.

Input/output analysis is a useful technique for sorting out and quantifying sector-by-sector impacts. This approach captures not only the direct effects but also related impacts in other parts of the economy. For example, an input/output matrix developed by the U.S. Department of Labor can estimate the impact of increased automobile sales on nearly 500 individual sectors in the U.S. economy.

### Transactions Matrix

Input/output analysis is based on a table that indicates historic patterns of purchases and sales between industries. Data for this table are usually generated from surveying a sample of firms. A simple two-sector version of such a table is shown in Table 5.6. The two sectors identified in the table are manufacturing and agriculture. Within each sector are firms producing specific products. For example, the manufacturing sector might include firms making electronic components and also firms that produce television sets and computers. Similarly, the agriculture sector could include firms producing cotton and wool.

The rows of Table 5.6 show the disposition of each sector's output. Consider first the manufacturing sector that is depicted in the first row. If the numbers represent billions of dollars, the first element in the first row indicates that firms engaged in manufacturing sold $8 billion of their product to other manufacturing firms. An example of these "within-sector" transactions might be the sale of electronic components to a television manufacturer. The second element of the first row shows that $10 billion of manufactured goods were sold to firms in the agricultural sector. Together, these first two elements indicate total sales by manufacturing to other firms. Such sales are sometimes called *intermediate sales* because they represent output that is used as inputs for making products sold to consumers. The first two elements of the second row are interpreted in a similar manner. They show intermediate sales by the agriculture sector.

The third element of each row is designated as final demand and shows sales to ultimate consumers. Finally, the last element is total sales. This number is the sum of the

**TABLE 5.6 Two-Sector Input/Output Table**

| | ***Sales to Manufacturing*** | + | ***Sales to Agriculture*** | + | ***Final Demand*** | = | ***Total Sales*** |
|---|---|---|---|---|---|---|---|
| Purchases from manufacturing | 8 | + | 10 | + | 2 | = | 20 |
| + | + | | + | | | | |
| Purchases from agriculture | 6 | + | 12 | + | 12 | = | 30 |
| + | + | | + | | | | |
| Value added | 6 | | 8 | | | | |
| = | = | | = | | | | |
| Total sales | 20 | | 30 | | | | |

**TABLE 5.7 Transactions Matrix**

| | *Sales to Manufacturing* | *Sales to Agriculture* |
|---|---|---|
| Purchases from manufacturing | 8 | 10 |
| Purchases from agriculture | 6 | 12 |

intermediate sales and final demand and represents the total sales by the sector. Thus the \$20 billion of total sales in manufacturing were made up of \$18 billion in intermediate sales to other firms and \$2 billion in sales to final consumers.

The columns of Table 5.6 indicate the use of revenues by each of the sectors. The elements in the first column show that firms in the manufacturing sector spent \$8 billion to purchase intermediate goods from other manufacturing firms and \$6 billion on goods from the agriculture sector. Subtracting that \$14 billion amount from total sector sales of \$20 billion leaves a residual of \$6 billion. This amount is referred to as value added because it represents the difference between the value of inputs purchased and the product finally sold. Value added consists of payments to workers, owners of capital, and the government.

The elements of the table indicating intermediate sales and purchases are of particular interest for input/output analysis. Together, they make up the transactions matrix of the model. Based on the numbers in Table 5.6, the transactions matrix is as shown in Table 5.7. In matrix notation, let this matrix be designated as $X$, where

$$X = \begin{bmatrix} 8 & 10 \\ 6 & 12 \end{bmatrix}$$

## Direct Requirements Matrix

If each element in a column of the transactions matrix is divided by the total sales of that sector's product, the result is the direct requirements matrix of an input/output model. The elements of each column of the direct requirements matrix can be interpreted as indicating the immediate or direct change in revenue for a sector generated by a \$1 change in demand for the product of the sector represented by the column. The direct requirements matrix for the sample problem is shown in Table 5.8.

Consider the elements of the manufacturing column. They were computed by dividing the first column of the transactions matrix by total sales in manufacturing—\$20 billion. The first element indicates that for each dollar of manufacturing sales, \$0.40 of intermediate sales to other firms in the manufacturing sector are initially generated. The second element shows that \$0.30 in sales by agricultural firms will be the direct result of each dollar of sales in the manufacturing sector. The elements of the second column have a similar interpretation with respect to each dollar of sales by the agricultural sector. Let the direct requirements matrix be designated as $A$. Thus

$$A = \begin{bmatrix} 0.40 & 0.33 \\ 0.30 & 0.40 \end{bmatrix}$$

**TABLE 5.8 Direct Requirements Matrix**

| | *Manufacturing* | *Agriculture* |
|---|---|---|
| Manufacturing sales per dollar | 0.40 | 0.33 |
| Agricultural sales per dollar | 0.30 | 0.40 |

**TABLE 5.9 Direct and Indirect Requirements Matrix**

| | *Manufacturing* | *Agriculture* |
|---|---|---|
| Change in total demand in manufacturing | 2.30 | 1.26 |
| Change in total demand in agriculture | 1.15 | 2.30 |

## Direct and Indirect Requirements Matrix

The direct requirements matrix shows the immediate or direct impact of changes in demand in a sector. But the elements of this matrix do not incorporate secondary and other impacts. For example, Table 5.8 indicates that the direct result of a $1 increase in agricultural sales would be a $0.33 increase in the demand for manufactured goods. However, such an increase would require that manufacturing firms purchase additional goods from other firms in the manufacturing and agricultural sectors. These purchases would stimulate still other purchases by firms in the two sectors. Thus the ultimate impact would be greater than predicted by the coefficient of the direct requirements matrix.

To take into account secondary impacts of changes in demand, input/output analysis requires computation of a direct and indirect requirements matrix. The derivation and calculation of this matrix are beyond the scope of this book, but conventionally it is designated by the notation $(I - A)^{-1}$, where $A$ is the direct requirements matrix, $I$ is the identity matrix, and the exponent denotes the inverse of a matrix.[4] For the example being considered, the direct and indirect requirements matrix is shown in Table 5.9.

The direct and indirect requirements matrix has a straightforward interpretation. The elements of each column indicate the change in total demand in each sector that would result from a $1 change in final demand for the sector designated by the column. For example, consider the first column. The first element shows that after all direct and secondary effects have been taken into account, a $1 change in final demand for manufactured goods will cause a $2.30 change in total demand in the manufacturing sector. The second element of the first column estimates that change in total demand in agriculture resulting from a $1 change in final demand for manufactured goods. In this case, the number is $1.15. Similarly, the elements of the second column show changes in total demand in each sector caused by a $1 change in final demand in the agricultural sector.

## Forecasting with an Input/Output Model

The direct and indirect requirements matrix can be used to forecast sector-by-sector impacts of changes in final demand. For example, suppose that an increase in exports causes a $5 billion increase in final demand in the agricultural sector. After taking into account all the direct and secondary effects, what impact would this change have on total demand for manufactured and agricultural goods?

With respect to manufacturing, Table 5.9 shows that each $1 change in agricultural final demand causes a $1.26 change in the total demand for manufactured goods. Thus

[4]An identity matrix has 1s on its main diagonal and 0s for its other elements. It is analogous to the number 1 in arithmetic. The inverse of a matrix is a matrix that if multiplied by the original yields the identity matrix. The concept is analogous to the reciprocal in arithmetic.

a \$5 billion increase would increase total demand in manufacturing by \$5 billion × 1.26, or \$6.30 billion. Similarly, the increase in total demand in the agriculture sector would be \$5 billion × 2.30, or \$11.50 billion.

Input/output analysis can also be used to forecast employment impacts. This is done by assuming a constant ratio of employment to total demand. To illustrate, assume that total demand in the manufacturing sector is \$500 billion and the total number of workers in that sector is 5,000,000. Hence the employment ratio is 1 to \$100,000. That is, one job exists for each \$100,000 in total demand.

If this ratio is constant, one additional job in the manufacturing sector will be created for each \$100,000 increase in total demand. Thus if a \$5 billion increase in final demand for agricultural products causes a \$6.3 billion increase in total demand in the manufacturing sector, the employment impact would be \$6.3 billion divided by \$100,000, or 63,000 new jobs. A similar computation could be made to estimate the employment increase in the agricultural sector.

## Example Automobile Sales and the Demand for Glass

Assume that by 2002, automobile sales to consumers are forecast to be \$10 billion greater than at present. A manufacturer of glass is contemplating a plant expansion and wants to know the probable effect of the increase in automobile demand on the total demand for glass. Union leaders in the industry want to know the employment impact.

The forecasts are to be based on a five-sector input/output model that includes sectors for glass and automobiles. The direct and indirect requirements matrix is as shown. The employment ratios are also shown. What will be the total demand and employment impacts in the glass industry of the projected increase in final demand for automobiles?

***Direct and Indirect Requirements Matrix***

| | *Steel* | *Automobiles* | *Computers* | *Glass* | *Electricity* |
|---|---|---|---|---|---|
| Steel | 1.034 | 0.334 | 0.008 | 0.010 | 0.010 |
| Automobiles | 0.008 | 1.010 | 0.009 | 0.002 | 0.007 |
| Computers | 0.003 | 0.004 | 1.110 | 0.001 | 0.100 |
| Glass | 0.009 | 0.090 | 0.048 | 1.004 | 0.005 |
| Electricity | 0.142 | 0.045 | 0.010 | 0.086 | 1.009 |

***Employment Ratios***

| | *Steel* | *Automobiles* | *Computers* | *Glass* | *Electricity* |
|---|---|---|---|---|---|
| Jobs per million dollars of total sales | 2.00 | 10.00 | 8.00 | 4.50 | 1.00 |

**Solution** The elements in the second column of the direct and indirect requirements matrix indicate the change in total demand resulting from a \$1 change in final demand for automobiles. For the glass industry, the coefficient is 0.090. Thus a \$10 billion increase in final demand for automobiles is estimated to increase total sales of glass by 0.090 × \$10 billion, or \$900 million. For the glass industry, the employment ratio is 4.5 jobs per \$1 million in total sales. Thus the \$900 million increase in total sales of glass is estimated to create 4,050 new jobs.

The construction of input/output models is a very expensive and time-consuming task. It would rarely be feasible for a firm to develop its own tables. Fortunately, several large models have been developed by various governmental agencies. Perhaps the most important is a 460-sector model of the U.S. economy maintained by the Department of Labor. The department also publishes employment ratios. This information is used by government analysts to evaluate impacts of public policy. It is also available for use by firms in forecasting. In addition, there are state and regional input/output models that can be used for forecasting on a more localized basis.

The major value of input/output analysis is that it takes into account the interrelationships between sectors. However, the approach has limitations in addition to its high cost. A fundamental problem is that input/output forecasts are based on ratios that are assumed to be fixed. This assumption may not be very realistic in situations where rapid technological change is occurring.

### Key Concepts

- Input/output analysis can be used to forecast the direct and indirect impacts of changes in one sector on other sectors of the economy.
- The transactions matrix shows the pattern of sales and purchases between sectors. The direct requirements matrix indicates the proportion of revenues used to purchase goods and services from each sector.
- The elements of the direct and indirect requirements matrix show the impact that a $1 change in final demand would have on total demand in each of the sectors of the model.
- Input/output analysis can be used to predict sector-by-sector sales and employment impacts.

## SUMMARY

Data for use in forecasting can be obtained from expert opinion, surveys, and market experiments. Forecasts of experts can be refined by procedures such as the Delphi technique, which uses an iterative process to aid individuals in assessing their judgments. Surveys of plans and attitudes by consumers and managers can be useful in forecasting. However, considerable care must be exercised to assure that the responses reflect subsequent consumer choices. Market experiments have an advantage over surveys because they provide information on actual consumer behavior. Their limitations include the risk of losing customers, high cost, problems of making consumers aware of changes, and inability to control all factors.

Variations in time-series data are the result of trends, seasonal fluctuations, cyclical patterns, and random factors. One method of analyzing a time series is trend projection. Basically, the technique involves estimating the coefficients that provide the best fit of data to a specified equation. The two most common equation forms estimate a constant rate of change and a constant percentage rate of change. Fitting the constant percentage rate-of-change equation requires that the data be inputted in logarithmic form. Forecasts based on trend projection of time-series data can be improved by adjusting for seasonal variations.

Another time-series technique for forecasting is exponential smoothing. Forecasts based on this approach are a weighted average of observed and predicted values for the last period. The weights are determined by choosing a smoothing constant that minimizes the sum of squared deviations between observed and forecasted values. Exponential smoothing works best when the time series does not have a strong time trend.

Barometric forecasting is based on time series that are correlated. If the changes in one series consistently occur prior to changes in another, a leading indicator has been identified. Useful leading indicators must be accurate, provide adequate and constant lead times, involve a logical explanation as to why one series predicts the other, and be available at reasonable cost. A composite index is a weighted average of several leading indicators. Diffusion indices measure the proportion of individual time series that increase from one period to the next.

Input/output analysis can be used to forecast the direct and indirect impacts of final demand changes in one sector on total demand in other sectors of the economy. The technique relies on data that show patterns of sales and purchases between sectors. These patterns are summarized in a transactions matrix. The direct requirements matrix is computed by dividing the columns of the transactions matrix by sector total sales. It shows the proportions of total revenues used to make purchases from the other sectors.

The direct and indirect requirements matrix is the basic tool used for forecasting with input/output analysis. Each number in a column shows the impact that a $1 change in final demand in one sector would have on total demand in each of the other sectors of the model. By assuming constant employment/total sales ratios, input/output analysis can also be used to estimate sector-by-sector employment impacts.

## Discussion Questions

5-1. The manager of Harrison Toys assembles a panel of five experts to assist in formulating a sales forecast. The intent is to use the Delphi technique, but after examining the projections of the others, none of the experts are persuaded to make any changes in their initial forecasts. What conclusion might the manager draw from the inability of the experts to agree on a projection?

5-2. A business surveys its customers to generate information about the demand for a proposed new product. Under what circumstances might the customers understate their true demand for the product? When might they overstate their true demand? Explain.

5-3. To test consumer response, on January 1, a soft-drink manufacturer introduces a new brand in the Chicago market area. Over a three-month period, sales are slow. and the company concludes that the product is unlikely to be profitable in nationwide marketing. Should the firm base its decision entirely on this experiment? Why or why not?

5-4. Using trend projection, how could a manager decide whether to use the constant-rate-of-change model rather than the constant-percentage rate of change?

5-5. Quarterly data from the beginning of 1988 to the end of 1998 are used to estimate the coefficients of the equation $S_t = S_0 + bt$, where $S_t$ is sales in the $t$th quarter. No seasonal adjustments are made. How accurate are each of the following forecasts likely to be? Explain.

a. Annual sales of a tire manufacturer for the year 2005.

b. Fourth-quarter sales for 1999 for a retail gift store.
c. Fourth-quarter sales for 1999 for a newspaper publisher.

5-6. A particular composite and a diffusion index are constructed based on the same three times series of data. In a given month, could the composite index increase while the diffusion index decreased? Use a simple example to explain your answer.

5-7. Could the lead time necessary for an indicator to be useful in forecasting be longer in one industry than in another? Give an example.

5-8. The average workweek is considered a leading indicator of general economic activity. Would longer workweeks predict an upturn or a downturn in economic activity? Explain.

5-9. A district manager proposes that her division develop its own state-level input/output model for use in forecasting. Is this a good idea? Why or why not?

5-10. A 20-sector input/output model gives value added as a percent of sales. Can this estimate be used to rank relative profitability of the sectors? Why or why not?

5-11. Would input/output analysis be useful in forecasting sales in a rapidly changing industry such as personal computers? Explain.

## Problems

5-1. The management of Romano's Foods decides to test-market a new frozen pizza. As the person in charge of the market experiment, what considerations should affect your decisions with respect to
a. The number and locations of cities in which the new pizza would be test-marketed?
b. The duration of the experiment?
c. Pricing of the product?

5-2. The trend line for sales of postcards (in millions) is estimated using annual data from 1968 to 1998. The result is $S_t = 10(1.05)^t$.
a. What is the interpretation of the "10" and the term in parentheses?
b. Forecast sales of postcards in 2000.

5-3. The trend for sales of beef (in millions of pounds) is estimated using data from the first quarter of 1972 to the fourth quarter of 1998. The result is as shown below with the *t*-value in parentheses:

$$S_t = 12.5 + \underset{(3.28)}{0.14t}$$

a. What is the interpretation of the coefficient of *t*? Is the coefficient statistically significant at the 95 percent level? Explain.
b. Forecast the demand for beef in the third quarter of 1999.

5-4. Quarterly sales for Swarthmore Cycles are as follows:

| *Period* | *Sales* | *Period* | *Sales* |
|---|---|---|---|
| 1997:I | 100 | 1998:I | 130 |
| 1997:II | 110 | 1998:II | 135 |
| 1997:III | 115 | 1998:III | 145 |
| 1997:IV | 125 | 1998:IV | 150 |

a. Plot the sales data and graphically fit a straight line to the points.

b. In using the data for trend prediction, would the constant rate of change or the constant percentage change model be more appropriate? Explain.
c. Using a calculator or the computation methods described on pages 115–116 of chapter 4, use the data to estimate the coefficients of the equation $S_t = S_0 + bt$, where $S_t$ is sales in the $t$th quarter.
d. Use the data to estimate the coefficients of the equation $S_t = S_0(1 + g)^t$.

5-5. Based on quarterly data from 1995:I to 1998:IV, Highland Foods estimates that potato chip sales can be projected using the equation

$$S_t = 5{,}000{,}000 + 100{,}000t$$

where 1995:I is period 1. Actual fourth-quarter sales were as follows:

| | |
|---|---|
| 1995 | 5,450,000 |
| 1996 | 5,860,000 |
| 1997 | 6,270,000 |
| 1998 | 6,680,000 |

a. Project sales for the first three quarters of 1999.
b. Without using a seasonal adjustment, project sales for 1999:IV.
c. Project seasonally adjusted sales for 1999:IV.

5-6. The winning heights (meters) for the Olympic pole vault for 1948–1996 were analyzed using the two trend-projection models. The estimated equations were

$$PV_t = 4.295 + 0.142t$$
$$PV_t = 4.289(1.033)t$$

where $PV_t$ is the forecasted winning height.
a. On average, how much has the winning height increased from one Olympics to the next? What has been the percentage increase?
b. Forecast the winning heights for 1976 and 1996.
c. The winning heights for 1976 and 1996 were 5.45 and 5.90 meters, respectively. Is there an explanation for the large differences between the forecasted and the actual values for 1996?

5-7 Over a six-year period, a firm's sales (millions) were \$200, \$220, \$180, \$200, \$190, and \$210.
a. Using a smoothing constant of 0.5, forecast sales for the next period.
b. Graph the original data. Based on the results from part (a), graph the forecasted sales. Do the smoothed results have less variability than the original data?
c. Forecast next period sales using smoothing constants of $\alpha = 1.0$ and $\alpha = 0.0$. What are the implicit assumptions associated with these smoothing constants?

5-8. For three periods, the average workweek is 40, 44, and 46 hours. An index of materials prices is 100, 110, and 105 for the same three periods. Consider workweek and materials prices as leading indicators of general economic activity.
a. Use these data to construct a composite index for the three periods.
b. What does the index predict regarding general economic activity for the fourth period? Explain.

5-9. The demand for newsprint over a 10-year period is shown in the following table. Also shown are three time series for the same time period.

| | Year | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Quantity of newsprint | 100 | 110 | 115 | 115 | 130 | 145 | 140 | 135 | 135 | 150 |
| Series *A* | 25 | 30 | 25 | 35 | 30 | 40 | 45 | 50 | 45 | 45 |
| Series *B* | 200 | 220 | 225 | 225 | 240 | 260 | 255 | 250 | 255 | 240 |
| Series *C* | 20 | 20 | 23 | 26 | 25 | 24 | 24 | 28 | 28 | 26 |

a. Which of the time series is a roughly coincident indicator of newsprint demand?
b. Which series is a leading indicator of newsprint demand? For the leading indicator, what is the lead time?
c. Based on the leading indicator, is newsprint demand likely to increase, decrease, or stay the same in period 11? Explain.

5-10. Each year, sector *A* purchases $5 million of goods from sector *B* and $10 million from sector *C*. Sector *B* makes annual purchases of $2 million from sector *A* and $8 million from sector *C*. Annual purchases of sector *C* are $3 million from sector *A* and $6 million from sector *B*. Within-sector purchases are $1 million for sector *A*, $4 million for sector *B*, and $5 million for sector *C*. Total sales in the sectors are $20 million, $20 million, and $30 million, respectively.
a. Write the transactions matrix for the three sectors.
b. How much is value added for sector *A*?
c. How much is final demand in sector *A*?
d. Write the direct requirements matrix for the model.

5-11. The Ohio Planning Commission has developed a four-sector regional input/output model with a direct and indirect requirements matrix as follows.

| | | Sector | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Sector | 1 | 1.20 | 0.45 | 0.90 | 0.60 |
| | 2 | 0.20 | 1.50 | 0.80 | 0.75 |
| | 3 | 0.89 | 0.63 | 1.10 | 0.10 |
| | 4 | 0.24 | 0.80 | 0.43 | 1.24 |

a. What is the interpretation of the element in the first row of the second column?
b. If final demand decreases by $5 million in sector 3, what will be the effect on total sales in sector 1?
c. The commission estimates that 1 ton of sulfur dioxide particles enters the atmosphere for each $10 million increase in total sales in sector 2. How much pollution will result from a $5 million increase in final demand in sector 1?

## Computer Problems

The following problems can be solved by using the TOOLS program (downloadable from www.prenhall.com/petersen) or by using other computer software.

5-12. Electric power consumption data from 1969 to 1983 were provided in problem 4-14 on page 141. Those data are to be used for this problem.
a. Using trend projection, estimate the constant rate of change in electricity consumption between 1969 and 1983.

b. Based on the result from part (a), forecast electricity consumption for 1984. Actual U.S. consumption was 777.4 billion kilowatt-hours. What might account for the large prediction error?

5-13. Monthly sales for a firm over an 11-month period are as shown.

| *Month* | *Sales* | *Month* | *Sales* |
|---|---|---|---|
| January | \$2,000 | July | \$1,550 |
| February | 1,350 | August | 1,300 |
| March | 1,950 | September | 2,200 |
| April | 1,975 | October | 2,775 |
| May | 3,100 | November | 2,350 |
| June | 1,750 | | |

a. For smoothing constants of 0.2, 0.4, 0.6, and 0.8, compute predicted sales. What are the sums of squared deviations? What is the optimal smoothing constant?
b. Based on the result from part (a), forecast sales for December.

5-14. The gold-medal-winning heights (meters) for the men's high jump for 1948 to 1992 are shown here.

| *Year* | *Height* | *Year* | *Height* |
|---|---|---|---|
| 1948 | 1.98 | 1972 | 2.23 |
| 1952 | 2.04 | 1976 | 2.25 |
| 1956 | 2.12 | 1980 | 2.36 |
| 1960 | 2.16 | 1984 | 2.35 |
| 1964 | 2.18 | 1988 | 2.38 |
| 1968 | 2.24 | 1992 | 2.34 |

a. Use trend projection to estimate the constant rate of change and the constant percentage rate of change for the period 1948 to 1992.
b. Use the estimated trend projection equations to forecast the winning height for 1996.
c. Use exponential smoothing with a smoothing constant of $\alpha = 0.2$ and $\alpha = 0.8$ to forecast the winning height for 1996.
d. Which approach (trend projection or exponential smoothing) provides the more accurate forecast? Explain.

5-15. Standard and Poor's provides a price index for stocks listed on the New York Stock Exchange. The index for 1972 to 1982 is shown here.

| *Year* | *Index* | *Year* | *Index* |
|---|---|---|---|
| 1972 | 109.2 | 1978 | 96.0 |
| 1973 | 107.4 | 1979 | 103.0 |
| 1974 | 82.9 | 1980 | 118.8 |
| 1975 | 86.2 | 1981 | 128.1 |
| 1976 | 102.0 | 1982 | 119.7 |
| 1977 | 98.2 | | |

a. What was the estimated constant percentage rate of change per period?
b. Using the data, project the Standard and Poor's Index for 1983 and 1984.

c. The actual index for 1983 and 1984 was

| *Year* | *Index* |
|---|---|
| 1983 | 160.4 |
| 1984 | 160.5 |

Is trend projection a good tool for estimating future stock prices? Why or why not?

5-16. Quarterly profits data (in millions of dollars) for 1996:I to 1998:IV are shown here.

| *Period* | *Profit* | *Period* | *Profit* |
|---|---|---|---|
| 1996:I | −1.0 | 1997:III | +1.8 |
| 1996:II | −0.5 | 1997:IV | +2.4 |
| 1996:III | −0.0 | 1998:I | +2.7 |
| 1996:IV | +0.6 | 1998:II | +2.9 |
| 1997:I | +1.1 | 1998:III | +3.3 |
| 1997:II | +1.3 | 1998:IV | +3.6 |

a. Using trend projection, what is the estimated constant rate of change per period?
b. What happens when the trend projection program is used to estimate the constant percentage rate of change? Explain.

5-17. Each year sector $A$ purchases \$5 million of goods from sector $B$ and \$10 million from sector $C$. Sector $B$ makes annual purchases of \$2 million from sector $A$ and \$8 million from sector $C$. Annual purchases of sector $C$ are \$3 million from sector $A$ and \$6 million from sector $B$. Within-sector purchases are \$1 million for sector $A$, \$4 million for sector $B$, and \$5 million for sector $C$. Total sales in each sector are \$20 million, \$20 million, and \$30 million, respectively. Final demands are \$14 million for sector $A$, \$5 million for sector $B$, and \$7 million for sector $C$.

a. Using input/output analysis, what are the elements of the direct and indirect requirements matrix? What is the interpretation of the element in the second row and third column?
b. Final demand in sector $A$ increases by \$2 million. What is the change in total demand for each sector?
c. Starting from the original values, final demand decreased by \$1 million each in sectors $A$ and $B$ and increases by \$2 million in sector $C$. What is the change in total demand for each sector?

5-18. Consider the following four-sector transactions matrix:

| | *Sales to A* | + | *Sales to B* | + | *Sales to C* | + | *Sales to D* | + | *Final Demand* | = | *Total Sales* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Purchases from $A$ | 8 | | 10 | | 5 | | 7 | | 10 | | 40 |
| Purchases from $B$ | 6 | | 9 | | 10 | | 8 | | 2 | | 35 |
| Purchases from $C$ | 10 | | 5 | | 2 | | 3 | | 15 | | 35 |
| Purchases from $D$ | 5 | | 8 | | 7 | | 5 | | 5 | | 30 |
| Value added | 11 | | 3 | | 11 | | 7 | | | | |
| Total sales | 40 | | 35 | | 35 | | 30 | | | | |

a. Write the direct and indirect requirements matrix. What is the interpretation of the element in the fourth column of the fourth row?
b. If final demand in sector $D$ increases by 5, what will be the changes in total demand for each sector?
c. It is estimated that 50 jobs are created for each one-unit increase in total demand in sector $A$ and 20 jobs for each one-unit increase in total demand in sector $D$. If final demand increased by 5 in sector $D$, how many additional jobs would be created in each of these two sectors?

# CASE
## Integrating Case Study II

## Southern Turkey

Southern Turkey's production division operates 10 turkey-producing facilities in Tennessee, Arkansas, and Kentucky. The firm's marketing division sells frozen turkey and turkey products in the southeastern United States. It has been in business since 1955 and currently has an 80 percent market share in the region. There are numerous other sellers, but they have little impact on Southern Turkey's operations. Quarterly sales data for the company (in thousands of pounds) for four years are shown in the following table.

**Quarterly Turkey Sales (thousands of pounds)**

| | *I* | *II* | *III* | *IV* |
|---|---|---|---|---|
| 1995 | 10,000 | 10,500 | 10,5000 | 14,000 |
| 1996 | 10,500 | 10,500 | 11,000 | 14,500 |
| 1997 | 11,000 | 11,500 | 11,500 | 15,000 |
| 1998 | 11,500 | 11,500 | 12,000 | 16,000 |

A concern for managers at Southern Turkey is that they lack information about the factors that affect demand for their product. In the past, some sales forecasts have been inaccurate because of fluctuations in the prices of substitute goods or changes in general economic conditions. To obtain needed data, management approved a marketing experiment in 100 cities in the firm's market area. This experiment was conducted during the first six months of 1997 and involved charging different prices for turkey and then measuring the per capita consumption in each city. In addition, data were collected on income per capita and prices of chicken, beef, and pork for each of the 100 cities. Then multiple-regression techniques were used to estimate the following per capita demand function for turkey:

$$\begin{aligned} Q = {} & 7.00 - 10.0\,P_T + 2.0\,P_C + 1.0\,P_B + 0.50\,P_P + .0003I \\ & (4.29)\quad (-5.642)\quad (3.333)\quad (2.876)\quad (0.503)\quad (2.000) \\ & R^2 = 0.75 \end{aligned}$$

where the $t$-statistics are in parentheses and

$Q$ = annual turkey consumption per capita (pounds)
$P_T$ = price of turkey
$P_C$ = price of chicken
$P_B$ = price of beef
$P_P$ = price of pork
$I$ = per capita income

At the present time, the average wholesale selling price of Southern's turkey is \$0.50 per pound. In the southeastern United States, the average prices of chicken, beef, and pork are \$0.50, \$1.50, and \$1.50 per pound, respectively. Income per capita in that region is \$18,000.

**Requirements**

1. How much turkey will the firm sell in 1999 and in 2000? How much will be sold in the fourth quarter of each of those years? (Recall that the fourth quarter includes the Thanksgiving and Christmas holidays.) Describe and justify your choice of a forecasting technique. Identify possible sources of error in your forecast.
2. At this time, the firm's production facilities are capable of producing 17,500,000 pounds of turkey every three months. However, management has tentative plans to expand capacity by constructing two new production facilities. These additions will be started in mid-2000 and be operational by June 2001. The expansion will increase the firm's production capacity to 19,000,000 pounds per quarter. Should management proceed with its timetable for expanding capacity? Why or why not?
3. Historically, prices have been determined by using a markup over production cost. Is there reason to believe that the current price of \$0.50 per pound is not the profit maximizing price? Explain.
4. In the data provided, does management have any information that would suggest how the demand for turkey is affected by general economic conditions? Explain. Be specific.
5. Government forecasts predict beef and chicken prices will be 10 percent higher in 1999, while pork prices will decline by 20 percent. The price of turkey is expected to be unchanged. Estimate the individual impacts of each of these changes on the per capita demand for turkey. How confident are you about your estimates? How would this change your forecast of sales made in part 1? Explain.