PART VII
Technological Change and Location Theory

CHAPTER 16

Technological Change in a Global Economy

CHAPTER 17

Locating the Firm in a Global Economy

INTEGRATING CASE

Study VII: GMG Entertainment, Inc.

CHAPTER 16

Technological Change in a Global Economy

PREVIEW

During the 1970s, Alvin Toffler authored an influential book called *Future Shock.*[1] A central thesis of this work is that change occurs so rapidly in modern society that people find it difficult to adapt to their evolving environment. He argued that the average person is emotionally and intellectually left behind by the rapid pace of technical and cultural change. If Toffler's ideas had any relevance for the 1970s, they are even more applicable as society moves into the next millennium. The last two decades brought radical changes in world political alignments, different societal values, and the availability of thousands of new and improved products because of advances in technology.

Faced with rapid change, many firms have found it difficult to keep pace. The demise of communism opened new markets in the former Soviet Union and Eastern Europe, but also reduced the demand for some products (such as military hardware) and intensified global competition as firms sought to take advantage of emerging opportunities. Complicating the situation is the need for firms to stay abreast of new developments within their own industry. Electronics companies must be competitive in using the new digital technologies for their audio and video equipment. Automobile manufacturers survive only if they keep costs down by using advanced robotics for assembly. Computer suppliers can stay profitable only if their machines include state-of-the-art chips, display terminals, and storage devices.

This chapter focuses on technology in today's global economy. The first section examines the impact of technological change. The second evaluates the relationship between the rate of technological change and alternative market structures. Section three considers characteristics of innovation and factors that result in successful innovation. The final section describes techniques that can be used to forecast future technological and societal changes.

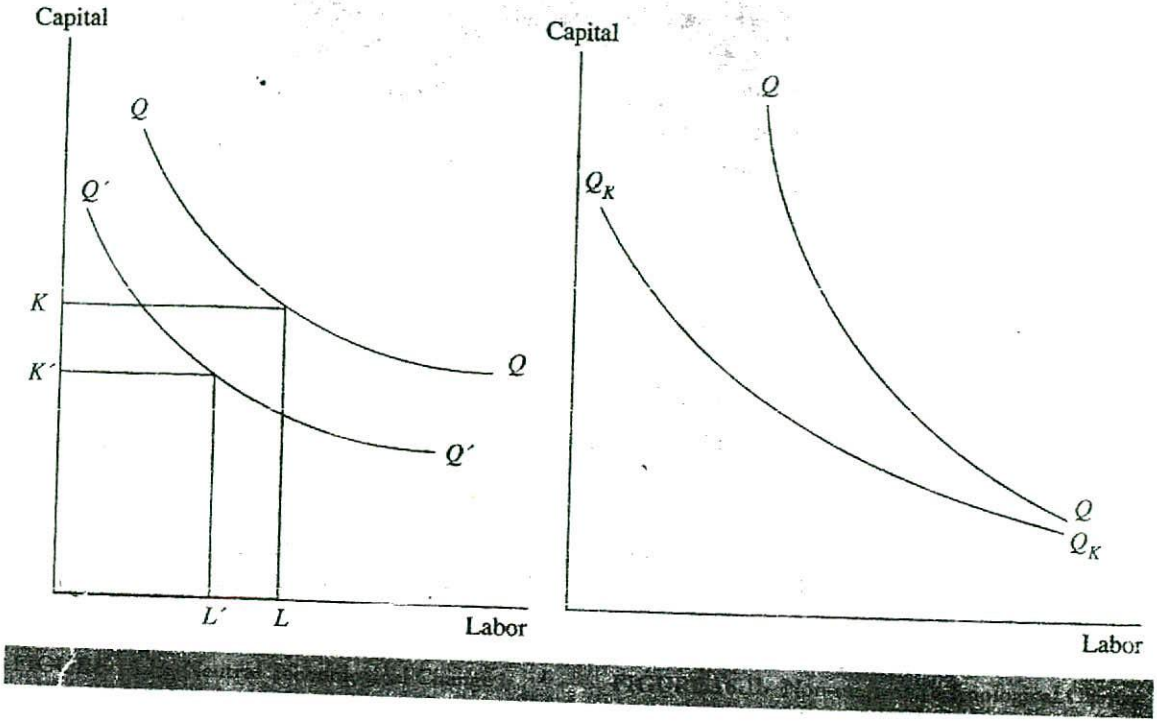
THE IMPACT OF TECHNOLOGICAL CHANGE

Technological change may involve new products, improvements or cost reductions for existing products, or better ways of managing the operations of a business. In some cases, the changes may seem simple and the results rather trivial, such as coating paper clips with colored plastic to prevent them from leaving marks on a page or tapering one side of the buttons on a shirt to make them easier to fasten.

In other cases, the technological advance may be brilliant and the impact on society highly significant. Consider the development and evolution of the electron microscope. The best optical microscopes can focus on objects as small as 1,000 angstroms in width.[2] During the 1930s, scientists learned to focus streams of electrons in the same way that optical devices focus light. The first electron microscopes achieved resolution of about 100 angstroms—ten times better than the optical microscopes. During the next 60 years, research efforts significantly improved the instruments. Today, commercially available electron microscopes can focus on objects as small as one angstrom—a thou-

[1] A. Toffler, *Future Shock* (New York: Random House, 1970).

[2] An angstrom is a unit of measurement equal to one ten-billionth of a meter.



sand times better than the best optical devices. This capability has allowed biologists, chemists, and physicists to make important discoveries. For example, it has enabled medical researchers to examine and manipulate bacteria, viruses, and genetic structures as they search for cures for diseases.

Technological Change and the Production Function

Technological change can be thought of as altering the firm's production function. Consider a product with an isoquant for 100 units of output, as shown by QQ in Figure 16.1a. This isoquant shows all the possible combinations of labor and capital that, if used efficiently, could produce 100 units. Often, technological change allows the firm to use fewer inputs. This possibility is illustrated by the 100-unit isoquant $Q'Q'$ in Figure 16.1a. Note that improvements in technology allow the same 100 units of output to be produced using less labor and capital. Technological change has caused a shift in the production isoquants.

The isoquants in Figure 16.1a suggest that improved technology is neutral; that is, it allows equal reductions in both inputs. This could occur, but it is usually not the case. Some technological advances are primarily laborsaving, while others are mainly capital saving. For example, the use of industrial robots in automobile manufacturing reduced the number of workers needed. In contrast, the development of the transistor conserved on capital by eliminating expensive and unreliable vacuum tubes and mechanical switches from television sets, radios, and other electronic equipment. Non-neutral

Table 16.1 Percentage Increases in Labor Productivity

Country	*Increase: 1970–1989*
United States	2.9%
Canada	2.5%
Denmark	3.3%
France	4.1%
Italy	4.9%
Japan	6.0%
Netherlands	4.9%
Sweden	2.9%
United Kingdom	3.8%
Germany	3.2%

technological change is illustrated by the isoquants in Figure 16.1b. The isoquant $Q_K Q_K$ depicts technological change that increased the marginal product of capital relative to the marginal product of labor.

Technological Change, Productivity, and Economic Growth

Economists use several measures to assess the performance of the economy. One of the most important is *productivity*, defined as the ratio of output to one or more inputs. Productivity is a key concept because it determines the standard of living that a country can achieve. In any given year, the total value of income received by individuals is based on the total value of goods and services produced. Thus, the only way for all consumers to have more real income is for the productivity of the inputs used to produce those goods and services to increase.

The most common productivity measure is *labor productivity*—output divided by the quantity of labor. If labor productivity can be increased, workers may be able to earn higher wages. Labor productivity has increased over time. However, there have been significant differences between countries. Percentage changes in labor productivity between 1970 and 1989 are provided in Table 16.1 for selected countries. Note that the rate of increase for the United States is less than every nation except Canada and Sweden. Slow growth in labor productivity has been a concern in the United States in recent years.

Technological change is an important source of increased labor productivity. By shifting production isoquants to the left, laborsaving technological change allows the same number of workers to produce more output. However, technological change is not the only component of increased labor productivity. Productivity increases as workers accumulate more human capital and as the capital stock of the economy increases. Changes in relative input prices can also affect the measured rate of labor productivity. For example, if capital becomes relatively more expensive, firms will use more labor and less capital. Thus, the ratio of output to labor input will decrease. Conversely, higher wage rates tend to reduce labor usage and to increase the measured rate of labor productivity.

TABLE 16.2 Sources of Growth in Real GNP

Source of Growth	*Percent Attributable to Source*
Increase in inputs	
Labor	20%
Education	19%
Capital	14%
Technological change	31%
Other factors*	16%
Total	100%

*Economies of scale, improved worker safety and health, pollution abatement, and fewer labor disputes.

The output–labor ratio is commonly used to measure productivity because it is easy to quantify. However, a better indicator is *total factor productivity,* which compares changes in output with changes in both labor and capital inputs. Using this approach, it is possible to identify the sources of economic growth over time. A study by Denison focused on economic growth as measured by the change in real income in the United States between 1929 and 1982.[3] The average annual growth rate in real income over the period was 2.8 percent. The estimated components of growth are shown in Table 16.2.

Note that the single most important source of economic growth between 1929 and 1982 was technological change. Almost one-third of the increase is attributable to improvements in technology. It is also important to observe that about 20 percent of growth was due to higher education levels of workers. Clearly, knowledge, whether associated with new or improved products or embodied in workers, is crucial to economic progress.

Key Concepts

- Technological change allows the firm to produce the same rate of output using fewer inputs.
- Technological change may be input-neutral, laborsaving, or capital saving.
- Labor productivity is the ratio of output to labor input.
- Total factor productivity compares changes in output to changes in both capital and labor.
- The single most important source of economic growth in the United States has been technological change.

[3]E. Denison, *Trends in American Economic Growth* (Washington: Brookings Institution, 1985).

Case Study

Labor Productivity: Automobile Production in England and Germany

In the early 1980s, the Ford Motor Company opened automobile plants in England and Germany. The two production facilities produced the same vehicle and, in fact, were identical. They had the same projected capacity and used the same robot welders, automated body presses, assembly lines, and other capital goods.

But when the plants began operation, the actual results were much different. The English plant produced 716 cars per day with 11,315 workers—an average of about 16 workers per car. In contrast, the German facility produced 1,027 cars per day using only 7,789 workers—an average of about 8 workers per car. Using identical equipment, labor productivity was more than twice as great in Germany as in England. What caused this dramatic difference?

Basically, it was the result of technical inefficiency in the English plant. When Ford officials began to compare the two facilities, they determined that there were differences in worker attitudes and behavior. Even a casual inspection of the plants revealed that the English employees weren't working as hard. Absenteeism was high and "goofing off" while on the job was frequently observed. In addition, featherbedding in the English plant had increased the number of workers. For example, a British doctor had certified that two men were required to lift the hood onto the car body, but typically one person did the lifting while the other stood by and watched.

Once they understood the problems, Ford management took steps to increase efficiency in the English plant. Employees were brought together in small groups and told that facility would be closed if productivity did not improve. British workers were flown to Germany to observe the operation of their sister plant. The firm also set up employee/management committees to assist in resolving conflicts.

By 1988, productivity at the English facility had markedly improved. Daily output increased to 1,130 cars with only 8,458 workers—matching the eight workers per car produced at the German plant earlier in the decade. However, productivity also increased in Germany. By 1988, that plant was using less than six workers per car per day. Thus, although significant improvement occurred in England, if the German facility is used as a standard, there was still technical inefficiency at the British plant. ■

TECHNOLOGICAL CHANGE AND MARKET STRUCTURE

The importance of technological change in facilitating economic growth is generally accepted. But an issue that has not been completely resolved is the relationship between alternative market structures and technological change. One question is, What type of market structure best facilitates the generation of new knowledge? Another is the direction of causality. Does the market structure determine the rate of technological change or does the nature of technology dictate which market structure will prevail?

The Effect of Market Structure on Technological Change

Some economists believe that market power is a necessary condition for rapid technological change. They argue that most modern research and development activities require huge investments and can take years before they yield results. Small firms operating in competitive markets may not have funds to allocate to such efforts and may be unable to take advantage of scale economies associated with complex R&D projects. Also, firms in competitive markets may be unable to capture all the economic profits resulting from their efforts. If competitors can easily imitate new products and product improvements, firms will be less likely to allocate their resources to R&D.

One of the most vocal advocates of this view is John Kenneth Galbraith. Using language reminiscent of Adam Smith's "invisible hand," Galbraith wrote:

> A benign Providence . . . has made the modern industry of a few large firms an almost perfect instrument for inducing technical change. . . . There is no more pleasant fiction than that technical change is the product of the matchless ingenuity of the small man forced by competition to employ his wits to better his neighbor. . . . Technical development has long since become the preserve of the scientist and the engineer.[4]

Clearly, large firms have been responsible for many important developments. AT&T's Bell Labs devised the transistor and Dupont introduced nylon. But small firms have also had an impact. Steven Jobs started Apple in his garage and revolutionized the computer industry. Photocopying was invented by a patent attorney, Herbert Carlson, and commercialized by a small firm that later became the Xerox Corporation. A study by Jewkes, Sawers, and Stillerman investigated the origins of 70 major inventions since 1880. They found that 54 percent of those inventions could be attributed to people working alone and another 11 percent involved individuals working with research institutions. Only about one-third of the inventions came out of industrial research laboratories.[5]

The case can be made that small firms in competitive markets might be more progressive than larger firms with monopoly power. Small firms may be more likely to provide an environment in which new ideas can flourish, while larger firms may impose bureaucratic rules that stifle creativity. Also, a young, small firm may have to be innovative to survive. In contrast, large, established firms that are partially insulated from competition may have very little motivation to change their product lines or production methods.

In evaluating the effect of market structure on technological change, the key is to consider both the ability and the incentives to be progressive. Ability involves being able to fund expensive R&D projects, withstand failures, and wait for results. Incentives include the need to remain competitive and being able to capture the rewards of technological advance.

Static versus Dynamic Inefficiency Large firms with market power have an advantage in facilitating technological change. They are also better positioned to capture the rewards. However, freedom from the need to compete may cause firms to be inefficient.

---

[4] J. K. Galbraith, *American Capitalism* (Boston: Mifflin, 1952), p. 91.

[5] J. Jewkes, D. Sawers, and R. Stillerman, *The Sources of Invention* (New York: Norton, 1969).

Year	Scenario I Output: Static Efficiency and 3% Growth in Factor Productivity	Scenario II Output: Static Inefficiency and 6% Growth in Factor Productivity
1	100.00	90.00
2	103.00	95.40
3	106.09	101.12
4	109.27	107.19
5	112.55	113.62
6	115.93	120.44

In chapter 9, it was argued that allocative inefficiency results when firms with market power set prices above marginal cost. Market power can also cause technical inefficiency if firms incur additional costs because they do not need to compete. Some economists believe that *static efficiency* (obtaining the maximum welfare from resources at each point in time) is less important than *dynamic efficiency* (increasing total factor productivity over time).

Assume that a monopoly continuously wastes 10 percent of its inputs because of static inefficiency. However, because the firm is heavily involved in research, total factor productivity increases at a rate of 6 percent per year. Also assume that if the firm operated in a more competitive market, static inefficiency would be eliminated but productivity increases would average only 3 percent per year. For a given quantity of resources, if the firm's maximum output is 100 units for the first year, output in subsequent years for each scenario is shown in Table 16.3. In the 5th year and beyond, the higher rate of growth in factor productivity overcomes the static inefficiency associated with market power and output is greater in scenario II.

Obviously, the numbers in Table 16.3 are hypothetical. If the static inefficiency was less, scenario II output would exceed scenario I output sooner. However, if the productivity differences were smaller, the inefficiency depicted by scenario II would take longer to overcome.

The issue of the optimal condition for promoting technological change has not been completely resolved. However, the theoretical arguments and empirical evidence suggest that no one market structure or firm size is clearly superior. Rather, diversity appears to be a virtue because each size and structure has its own advantages and disadvantages.

R&D and the Prisoner's Dilemma In addition to traditional static and dynamic inefficiency, there are situations where research and development efforts can be wasteful. Consider the case of firms in a duopoly. Each firm has the option of being involved in R&D to find ways to reduce production costs or just using its existing technology and saving the R&D expense. If one firm cuts costs and the other does not, the more efficient firm will have a competitive advantage. But if both firms find ways to reduce costs, assume that competition between them causes savings to be passed on to consumers and the firms receive little benefit. The payoffs to R&D efforts are shown in Table 16.4.

TABLE 16.4 Research and Development Payoffs

		Firm B	
		R&D	***No R&D***
Firm A	***R&D***	80, 40	160, −40
	No R&D	−40, 120	120, 80

Table 16.4 is a classical example of the prisoner's dilemma discussed in chapter 11. If both firms choose the R&D option, their expenditures will be partially self-cancelling. But if one firm is involved and the other is not, the non-R&D firm will lose money. Thus, each firm will engage in R&D. But both firms could earn more profit if they could agree not to spend for R&D. Unfortunately, this is not very likely. It is more difficult to monitor a competitor's R&D efforts than other activities such as pricing and advertising. Also, unlike prices and advertising, successful R&D efforts cannot be quickly and easily matched.

The prisoner's dilemma characterizes R&D efforts in many concentrated industries in the United States. Firms spend huge sums of money on research activities that duplicate those of competitors. Antitrust laws prevent them from engaging in joint research efforts, and the fear of falling behind makes it impossible for them to cut back. In contrast, in Japan the federal government encourages joint R&D and large firms are much more likely to share information with each other about new products and processes.

The Effect of Technological Change on Market Structure

Market structure can affect the rate of technological change, but technology can also significantly affect the structure of the market. Telecommunications is a good example. The basic telephone patent was issued to Alexander Graham Bell in 1876. This patent right provided the Bell System a monopoly until it expired in 1893. For a few years, there was vigorous competition in some cities, but the available technology soon caused the industry to evolve into a virtual monopoly. In each city, the local network required that lines be deployed to every home and business in the community. This involved obtaining rights of way, putting wires under streets, and erecting thousands of telephone poles. Establishing a nationwide long-distance network required that tens of thousands of miles of copper wire be run over great distances and through often difficult terrain. The necessary capital equipment was extremely expensive and there were significant economies of scale associated with the endeavor. As a result, smaller firms were unable to compete, and AT&T was the dominant firm for almost one hundred years.

In the late 1960s, development of microwave technology provided opportunities for competitors. As an alternative to running cable, firms could establish communication links by setting up microwave towers every twenty or thirty miles. Microwave Communications, Inc. (MCI) offered the first serious challenge to AT&T by offering data transmission service between St. Louis and Chicago. After ten years in court and numerous challenges to the Federal Communications Commission, MCI finally established itself as a viable competitor. More recently, satellites and fiber optics have emerged as alternative technologies. Today, AT&T still dominates the long-distance market, but it faces significant competition from firms such as MCI and Sprint.

At the local level, technological change is having a similar effect on telecommunications market structure. Until the mid-1980s, most industry analysts believed that local telephone networks were natural monopolies and that regulation would always be necessary. But in the last few years there have been some dramatic changes. Large businesses have begun to establish their own microwave, satellite, and fiber optic links to meet their communication needs. Cellular systems provide an alternative to using the Bell networks. Potentially most important, many cable companies are installing new technology that allows two-way communications. In the near future, it is very likely that cable and telephone companies will be competing with one another. The cable firms will offer telephone service and the telephone companies may offer movies, games, and shopping services to their customers. Another scenerio is that these firms will merge and provide a broad range of services.

Some casual observers of the industry have suggested that the recent dramatic changes in telecommunications market structure are primarily the result of changes in regulatory policy, as the courts and Federal Communications Commission have permitted competition in markets where monopolies had been protected. But a closer examination reveals that government was simply responding to the forces of technological change. When telecommunications technology no longer dictated the existence of natural monopoly, it was impossible for bureaucrats to perpetuate this structure. The new technologies made increased competition inevitable.

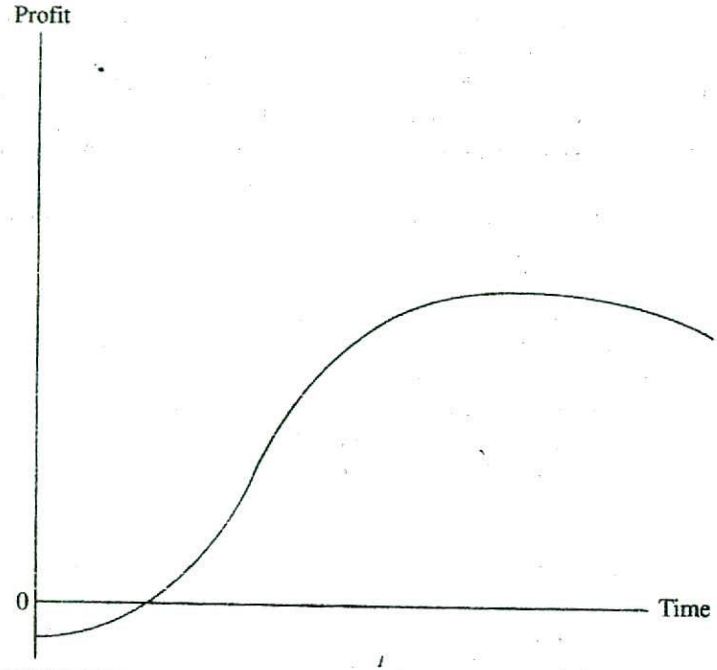
The same forces have caused structural change in the computer industry. Technological change has resulted in faster computer chips and storage devices. This has changed the market because many applications that previously required a large mainframe computer can now be performed on personal computers. The result is that the industry has evolved from an oligopoly dominated by a few firms such as IBM into a competitive market with hundreds of firms selling personal computers. As with telecommunications, the new market structure reflects changes in technology.

Key Concepts

- Large firms with market power may be more innovative because they can afford the large investments, take advantage of scale economies, and capture the rewards from their efforts.
- Arguments that small firms will be more progressive include the need to innovate to survive and lack of bureaucratic constraints on creative activity.
- R&D efforts may be wasted in oligopolies if firms engage in efforts that duplicate the activities of other firms.
- Changes in technology, by reducing or increasing economies of scale, can alter market structures.

INDUSTRIAL INNOVATION

To this point, the term *technological change* has been used rather loosely. Sometimes it can be useful to distinguish among invention, innovation, and diffusion. *Invention* can be thought of as the creation of new ideas. *Innovation* represents taking those ideas and transforming them into something that is useful for society. *Diffusion* is the process whereby the new product or process becomes available throughout the society.



In many cases, inventions never get to the innovation stage. In other cases, innovations fail to become widely adopted. Sometimes the problem is that an invention or innovation provides no real technical advantage or is not economically viable. But there are also instances when truly beneficial ideas languish for many years. In the early days of sailing, scurvy was the worst killer of sailors. In 1601, an English sea captain found that two or three teaspoons of lemon juice a day provided almost complete protection against the disease. His finding was quite well known at the time, but it was not until 1865, over two hundred and fifty years later, that the remedy was widely used and scurvy ceased to be a threat among British sailors.

Product versus Process Innovations

Innovation can be divided into two broad categories. *Product innovation* involves the bringing of new goods or services to the market, while *process innovation* is concerned with new techniques that reduce costs of producing and distributing existing products.

For a successful product innovation, the good or service will generate a stream of economic profit, such as that shown in Figure 16.2. In the early years, the firm may lose money as it attempts to launch the product and gain consumer acceptance. Later, there may be a period of rapid growth as the good or service becomes widely used and competitors provide substitutes. After the product has been available for a number of years, sales may stagnate or even decline.

In evaluating a product innovation, the firm can use the techniques of capital budgeting. An innovation usually involves a substantial initial cost and a stream of future profits. Basically, the question is whether the present value of profits is likely to exceed the up-front cost of bringing the product to the market.

Evaluation of process innovation is similar. The initial cost of implementing the innovation must be balanced against the future cost savings that will result from the improved process. If the net present value is positive, then the new technique should be incorporated into the production process.

To illustrate the firm's decision, consider the example of a new process that could reduce the fuel cost of producing electricity. Suppose that an electric utility has generators that have been in service for many years. At present, there is no capital cost—the total cost of generation is the cost of fuel plus the firm's operation and maintenance expense. If the new generating technology is adopted, the firm could reduce its operating costs but would have to pay the initial capital cost. For the process innovation to be profitable, the operating cost savings would have to be greater than the initial cost.

Assume that in deciding to use the new generators, managers consider costs for the next T years. Also assume that there is no inflation and that the equipment will produce Q kilowatt-hours of electricity each year. The operating cost per kilowatt-hour with the new equipment is estimated to be OC_N and the operating cost using the old generators is projected to be OC_E. Because the new generators are more efficient (i.e., use less fuel and will have lower maintenance costs), it is known that $OC_E > OC_N$. However, in calculating the operating cost saving, the time value of money must be considered—a dollar saved 10 years from now is less valuable than a dollar saved today. This is taken into account by discounting the cost savings.

For the existing equipment, the present value of operating costs ($PVOC_E$) is given by

$$PVOC_E = Q \cdot \sum_{t=1}^{T} \left[\frac{OC_E}{(1 + r)^t} \right]$$

where r is the discount rate. The present value of operating costs for the new generating process ($PVOC_N$) is given by

$$PVOC_N = Q \cdot \sum_{t=1}^{T} \left[\frac{OC_N}{(1 + r)^t} \right]$$

Hence, the present value of the operating cost saving ($PVOC_S$) is

$$PVOC_S = Q(OC_E - OC_N) \sum_{t=1}^{T} \left[\frac{1}{(1 + r)^t} \right]$$

Let the initial cost of the new generating technology be IC and the present value of the salvage value of the old equipment be SV. The innovation decision should be based on a comparison of the purchase cost less salvage value versus the present value of the cost saving. Specifically, if $IC - SV < PVOC_S$, the firm should adopt the innovation. Otherwise, it should continue to use the existing technology.

Case Study

Successful Innovation: The Gillette Sensor Razor

Gillette's Sensor razor is an example of successful product and process innovation. In the mid-1980s, engineers at Gillette perfected the design of a new hand razor with flexible twin blades that adjusted to the contours of the shaver's face. The new design provided a closer and smoother shave than any razor on the market.

When it was introduced in 1990, the Sensor was a near instant success. In the first two years, the company sold 50 million razors and more than two billion of the twin-blade cartridges. Today, the product is by far the top-selling razor in the United States, with over 40 percent of the market for nondisposable razors—nearly triple that of its nearest competitor.

The design of the new product was the easy part of the Sensor project. The challenging problem was finding technology that would allow the firm to profitably manufacture millions of razors and blade cartridges each year. With that volume, a reduction of even a tenth of a cent per unit could mean a large increase in profit.

One of the problems was that a Sensor blade cartridge has 10 parts, while other twin blades have no more than 6. More importantly, the blades must be mounted on a spring that permits them to move up and down but not laterally. Assembling the product at high rates of speed with margins of error between parts as small as one-thousandth of an inch was a difficult challenge. Gillette succeeded by using high-technology laser welding equipment that can precisely complete the necessary 15 welds on each cartridge in less than one-fifth of a second. Quality control is also an important part of the manufacturing process. Each Sensor cartridge undergoes 100 mechanical or electronic inspections before it leaves the assembly line.

The firm's use of state-of-the-art technology and attention to detail has paid off in cost control. In the first 2 years since it started making Sensor, Gillette cut unit costs by 30 percent and hopes to reduce costs by another 10 percent over the next 2 years. Even the smallest detail can make a difference. A product inspection team noticed that a very small percentage of defects in a clip used to anchor the cartridge assembly was causing periodic slowdowns on the assembly line. The problem was detected only after analyzing the data from millions of assemblies. By making a slight design change, the firm was able to increase its rate of output by four percent with no increase in costs.

The Sensor's success illustrates the connection between product and process innovation. The flexible blade design was a brilliant design concept, but it became profitable to Gillette only after a series of process innovations reduced costs. ■

Requirements for Successful Innovation

Often, the firm that initially introduces a new product does not reap the rewards. For example, Bowmar Instruments was the first to develop pocket calculators, but the firm went out of business because it was unable to withstand competition from Texas

Instruments, Hewlett Packard, and other large firms. RC Cola was the first company to sell cola drinks in a can and also the first to introduce a diet cola, but Coca-Cola and Pepsi-Cola responded so quickly that the smaller firm never gained any significant advantage from its innovation. In the early 1970s, Electrical Musical Industries, a small English firm, first developed CAT-scan technology, which provides a cross-sectional view of the internal organs of the human body. The technique is probably the most important advance in radiology since the discovery of the X-ray in 1895. But the firm had little marketing capability and no involvement in medical electronics in the U.S. The result was that firms such as General Electric soon introduced their own scanners, and Electrical Musical Industries was forced out of the business in less than 10 years, after losing more than $50 million.

The key to successful innovation is the ability to capture the rewards of new product or process developments. This can be difficult because technology, which is basically just information, has some of the characteristics of a public good. Once a new idea has been conceived, many people can use the information simultaneously. For the information to be valuable to its creator, there must be some way to prevent or at least delay others from gaining access to it.

Patents can provide some protection, but frequently they can be "invented around" and are usually not very effective for process innovations. In general, the keys to capturing the rewards of new process innovation are secrecy and being first to use the process. The advantage of being first is that the innovator can accumulate experience, which allows it to keep costs below those of other firms. For new product innovation, the keys to success often are the ability to market the product and, depending on the nature of the good, provide high-quality service to consumers.

Product innovation typically evolves through several stages. In the early period, the first firm to the market may have a temporary monopoly. As imitators appear, there may be several competing designs available. Over time, designs that are more costly to produce or less attractive to consumers will be dropped and the market will enter an era of product standardization. In the early days of the automobile industry, firms produced vehicles with both steam and internal combustion engines. History records that the internal combustion design was the winner. When videocassette recorders were first introduced, there were two competing technologies—Beta and VHS. The Beta technology actually had advantages but never gained a large market share and now has been abandoned.

Once a (more or less) standard design for a product has been determined, the next phase involves developing better manufacturing processes that can reduce costs. At this point, economies of scale and scope can become important. Firms able to produce at high rates of output may be able to eliminate their small, less efficient competitors. Firms with experience in manufacturing similar products may be able to transfer their resources and expertise to producing the new good. Typically, a firm has a set of core competencies or areas where it has advantages over other firms. Innovation is more likely to be successful when the firm stays with these core competencies. Although there are exceptions, companies that attempt to innovate in areas unrelated to their previous business experience often fail. The case of Electrical Musical Industries cited earlier in the chapter is a good example. The firm simply did not have the needed background to become involved in the health care field.

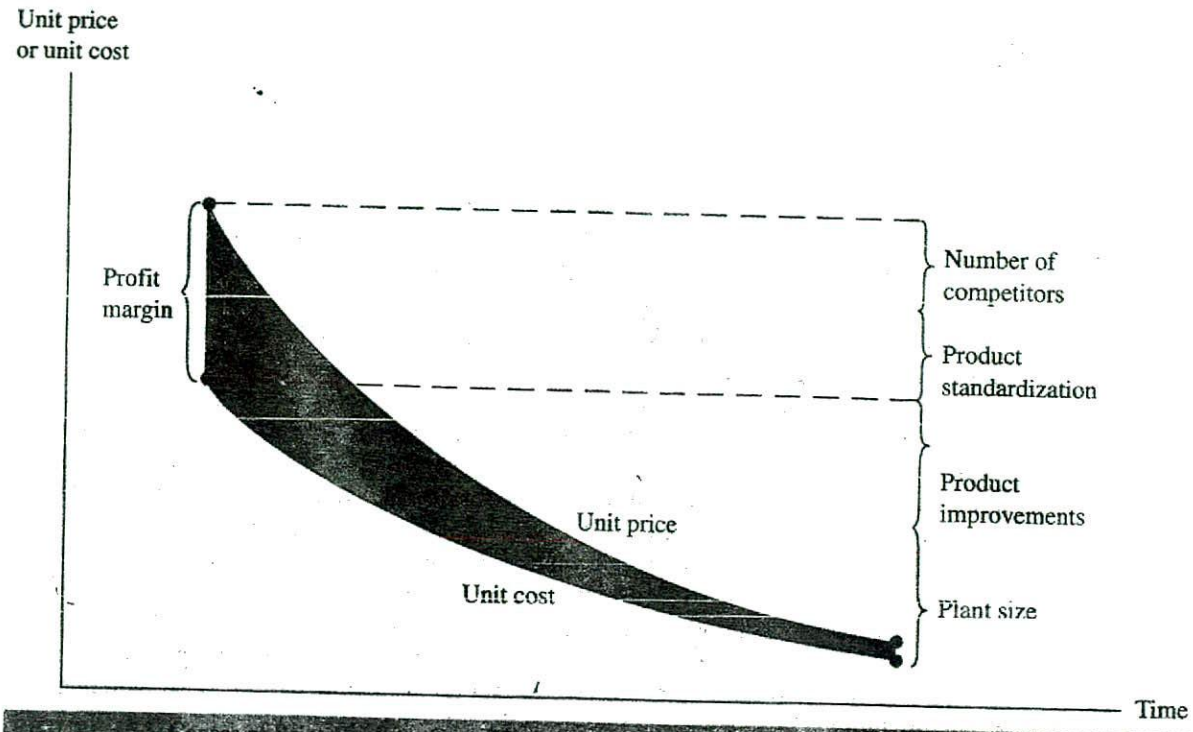


Figure 16.3 summarizes the primary sources of price declines over time for new products. The main causes of cost reductions are larger plant sizes and improvements in the production process. But profit margins also tend to decrease because product standardization eliminates the product differentiation advantage of early innovators and because of increased competition resulting from the entry of new firms into the market.

Key Concepts

- Innovation involves developing useful products. Diffusion is the process whereby new products become widely available.
- Product and process innovations should be evaluated by comparing the initial cost to the present value of the expected incremental profit.
- Often, the keys to successful product innovation can be marketing and the ability to provide service. Profitable process innovation involves secrecy and being able to exploit economies of scale and scope.
- Innovation is more likely to be successful if the firm stays within its core competencies.

Case Study
The Qwerty Dilemma

Until the late 1800s, there was no standard arrangement of the keys on a typewriter. In 1873 Christopher Scholes developed the QWERTY layout, named after the six letters in the top left row. Ironically, the QWERTY keyboard was designed to increase rather than reduce typing time. During that era, all typewriters were mechanical and keys tended to jam if they were struck too rapidly. The placement of letters on the QWERTY keyboard made typists do more reaching and slowed their speed.

With today's electronic typewriters and word processors, the jamming problem no longer exists. The upper limit of speed is now constrained by human and not mechanical factors. In 1932, August Dvorak used time and motion studies to develop a new keyboard layout, which allowed faster typing. The Dvorak keyboard was set up so that the fingers rested on the letters, A,O, E, U, I, D, H, T, N, and S. Letters that were used less frequently were placed on other rows. The result was that about 70 percent of typing was done on the home row. In contrast, the QWERTY keyboard uses the keys on which the fingers rest for about 30 percent of strokes. The Dvorak board even takes into account the differential strength of fingers by placing keys so that the stronger fingers are used more often. It also is engineered so that successive key strokes use alternate hands. While one hand is hitting a key, a finger on the other hand can be reaching to strike the next key. It is generally agreed that the Dvorak keyboard can reduce typing time by as much as 10 percent, and that it also reduces hand fatigue.

Because of its clear superiority, the Dvorak keyboard should have become the dominant design. But it didn't work out that way. In fact, it is difficult to find a Dvorak typewriter. The problem is that the QWERTY board has been the industry standard for many years. The existing base of equipment is almost entirely QWERTY, and few typists know the Dvorak system. Because there is little demand for the improved keyboards, manufacturers do not produce them. But, because they are not available, typists have little opportunity to retrain. The result is a vicious circle that perpetuates the use of inferior technology. ■

Patents and Innovation

Advocates of a patent system usually make their case on the argument that market forces may fail to provide sufficient economic incentives to reward innovation. If imitators are able to move in rapidly and capture a substantial share of the market, the initial profits earned by innovators may not provide adequate compensation for the costs and risks they bear. However, if there is a substantial delay between the time of innovation and successful entry by competitors, the economic profits earned in the interim

may make invention and innovation a more attractive activity. The patent system, by establishing a period of time during which the firm faces reduced competition, increases the expected return for innovative effort.

By stimulating technological change, the patent system can increase the flow of new products and processes to the market. The case for patents rests on a benefit/cost analysis. The assumption is that new products and processes available because of the patent incentive more than compensate society for the higher prices that temporarily result from the monopoly status given patent holders.

In the United States, patents confer the exclusive right to the use of an idea for a period of 17 years. Under U.S. law, "any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof" can be patented. However, there are three criteria that must be satisfied to obtain a patent. First, the invention must be new. Specifically, it must not have been known to the public before the inventor completed it or for more than 1 year prior to a patent application. Second, it must be useful. In practice, this test can be satisfied if there is at least some indication that the idea can be put to a practical use. Finally, it must be nonobvious. This provision has been controversial, but the present standard is that an invention cannot be patented if "the subject matter as a whole would have been obvious at the time the invention was made to a person with ordinary skill in the art."

A secondary reason for granting patents is to provide for widespread disclosure of new ideas and techniques. One of the requirements to obtain a patent in most countries is that the applicant must describe his or her invention in sufficient detail to permit others "skilled in the art" to use it if they had permission. A firm must take care submitting a patent application. The claims must provide enough information to satisfy the patent examiner but not enough to allow potential competitors to invent around the patent. The document must also be sufficiently broad to give the company latitude in marketing the product or process but narrow enough to avoid being rejected because it conflicts with the claims of existing patents.

Once a new patent is granted, its value may be tested in the courts. The U.S. Patent Office grants the patent but provides no other legal assistance beyond that point. If the owner of a patent believes that someone has infringed on a patent right, to get redress, the person or firm must sue the other party. An interesting recent example is the long legal battle that inventor Robert Kerns waged against the automobile industry. Kerns held the patent for a windshield wiper system used on almost all U.S. cars and trucks. He charged that the automobile industry was illegally using his invention without paying him royalties. After 14 years, in 1992, the courts ordered Ford to give him $10.2 million and Chrysler to pay $11.3 million in compensation. Not all infringement decisions favor the patent holder. In fact, about 20 percent of all court cases involving patents ultimately result in the patents being declared invalid.

One of the difficult aspects of U.S. patent law is the principle that a patent is issued to the person who conceives the idea rather than the person who first files for a patent. Although this may seem fair, it complicates the issuance of patents because the patent office (and often the courts) must decide who was the original inventor. The United States is alone in using this criterion. Every other country awards patents based on who is the first to file an application. In the next few years, it is likely that the United States will conform and adopt the first-to-file principle.

A second important international issue involving patents and also copyrights is countries that allow firms to steal and copy protected ideas. Computer software and hardware companies, pharmaceutical firms, book publishers, watchmakers, and shoe manufacturers in the United States and other countries lose billions of dollars each year because imitations of their products are sold to unsuspecting consumers. Among the worst offenders are firms in Korea, Taiwan, India, and Thailand. Either because of a lack of interest or resources for enforcement, governments of these nations seldom prosecute local firms for patent and copyright violations.

Key Concepts

- By creating a lag between innovation and imitation, patents increase the incentive for innovation.
- U.S. patents have a duration of 17 years.
- To be patented, an invention must be new, useful, and nonobvious.
- The award of a patent does not guarantee that the patent holder's rights will be upheld in court.
- U.S. patents are awarded based on who is first to conceive of an idea, while other nations use a first-to-file standard.

TECHNOLOGICAL AND ENVIRONMENTAL FORECASTING

One of the requirements for effective long-term planning by managers is to assess the changes in technology and environmental conditions that could affect the firm. Technological forecasting involves anticipating development of new products and processes and the time it will take for them to be widely adopted. Environmental forecasts focus on factors such as population growth, resources, and social and political trends that may affect the firm's future.

All predictive activity is subject to error, but technological and environmental forecasting is particularly difficult because it often involves assessing ideas and relationships that do not exist at the time the analysis is being performed. For example, nuclear fusion is a technology for energy generation based on the joining of atomic particles. For the last thirty years, scientists have been predicting that commercial fusion would soon be available. But the fusion researchers are still in their laboratories and the breakthroughs have not yet materialized.

Although the techniques discussed in chapter 5 have some value for long-term forecasting, most are better suited for predicting performance a year or two in the future. Long-term technological and environmental forecasts require special methods. By their very nature, they are less precise and rely heavily on the skill of the analyst. Two broad categories of techniques are discussed here. The qualitative methods focus on subjective judgments of individuals, while the quantitative approaches attempt to manipulate the limited data that are available.

Qualitative Forecasting Methods

When there are no empirical data, technological and environmental forecasts must be based on the best guesses of people who are knowledgeable in the field. In chapter 5, the Delphi method was introduced as a technique for refining expert opinion. To use this approach, a group of experts is asked to assess a particular situation, be presented with the judgments of others in the group, and then to reevaluate their individual positions based on what they have heard. This process may continue through several iterations until a consensus is reached or until it is apparent that there will be no consensus. The Delphi method has been successfully used to forecast the nature and timing of technological change. Gerstenfeld describes the following application.[6]

Round 1. A panel was requested to list inventions and technological changes that they thought were both needed and achievable in the next 50 years. From their responses, a list of 49 items was prepared.

Round 2. The group was presented with a list of time periods and asked to forecast the interval in which there was a 50 percent chance that each of the 49 developments would first be available. For example, one panel member might respond that there was a 50 percent chance of a cure for all forms of cancer 11 to 20 years in the future. Another might choose a 21- to 30-year interval.

Round 3. On items for which there was no consensus, the experts were asked to indicate the reasons for their estimates. After assessing the views of others, they were given an opportunity to change their responses.

Round 4. The process of rounds 2 and 3 was repeated. The ultimate result was a narrower range of estimates for most breakthroughs.

The Delphi method is not the only way to assist experts in making subjective assessments of the future. Sandoz, a Swiss manufacturer of pharmaceuticals, used subjective probabilities as a basis for R&D planning decisions. Twice each year a small group of managers and scientists was asked to estimate the probability of success for each of the firm's R&D projects. One of the problems the firm experienced was that the group had difficulty in making numerical estimates of the probabilities.

To assist the experts, an indirect method of quantifying probabilities was used. A wheel was divided into two colored sections, one blue and one orange. The wheel was adjustable so that the relative proportions of the two colors could be changed. In the center was a pointer that could be spun. Each expert was asked which event was more likely: (1) that the spinner would stop on the orange section or (2) that a specified R&D project would succeed. If the answer was (1), the wheel was adjusted to decrease the relative size of the orange. If the answer was (2), the orange portion of the wheel was made larger. The procedure was repeated until the person believed that the two events were equally likely. The portion of the wheel that was orange was the expert's subjective probability that the R&D project would be successful. The probability wheel provided no new information, but it did give the scientists and managers a frame of reference for making their decisions.

---

[6]A. Gerstenfeld. "Technological Forecasting," *Journal of Business* (January 1971): 10–18.

Forecasts of future technology can also be made using analogy methods. The concept is that experience with existing technology may provide insights in predicting what will happen with a new product or process. One possible application involves forecasting new technology in commercial aviation. Historically, the most advanced aircraft have been military because the perceived needs of defense have resulted in billions of dollars being spent for R&D. However, many of the resulting innovations have later been incorporated into commercial aircraft. Thus, by tracking the trends in military aviation, it may be possible to forecast future developments in the commercial aviation area.

The analogy method can be used if there are technologies that are expected to follow similar paths. The use of solar heating has been analyzed by examining historical experience with heat pumps, and the diffusion of color television sets was predicted using sales trends for black and white sets. Similarly, the future rate of adoption of high-definition television sets could be evaluated based on experience with large-screen TV sets.

Case Study
Megatrends 2000

In 1982, John Naisbitt wrote a popular and influential book, *Megatrends,* in which he predicted 10 social trends that would affect the future. His analysis was based on a method called content analysis, which was developed during World War II. During the war, intelligence personnel wanted to obtain information on trends and opinions in enemy nations. In the United States, such information could be collected by developing and administering an opinion poll, but that option was not very feasible in Germany at the time. The best alternative was to perform an in-depth analysis of items appearing in German newspapers. As the war continued, information about the economy and public morale was pieced together and used to evaluate conditions and trends in Germany. The project was so successful that it was also used by the U.S. government to analyze Japan, Korea, and Vietnam during wars with those countries.

Naisbitt and his associates use this same technique to forecast the directions in which U.S. society is heading. For years, his staff extracted information from big-city and small-town newspapers. Relevant articles were clipped and assigned to one or more categories. By analyzing developments and changes found in their files, they attempted to identify important trends. Firms and other organizations subscribed to this service and received forecasts pertaining to a specific area.

In 1990, Naisbitt published a second best-selling book called *Megatrends 2000.* His approach was the same, but this time the 10 most important trends for the 1990s were identified. Following are his predictions.

1. A global economic boom during the 1990s
2. A renaissance in the arts
3. The emergence of free-market socialism
4. A trend toward global lifestyles, but at the same time, intense nationalism
5. The privatization of state-owned firms
6. Increasing importance of the Pacific Rim countries

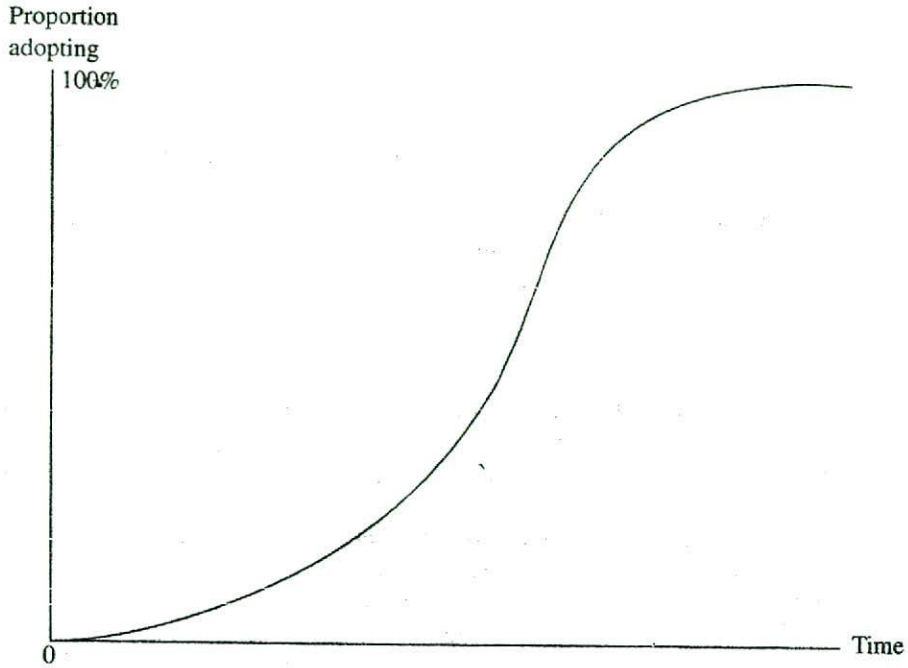


FIGURE 16.5 The Diffusion Curve

Iowa farmers and found that all but two had switched to the new seed between 1928 and 1941. From 1928 to 1933, the adoption rate was only 10 percent. But during the next 3 years, an additional 40 percent made the change. Thereafter, the rate of adoption began to level off because fewer farmers remained to adopt the new idea. Ryan and Gross were pioneers, but their methods have been used many times since then to evaluate the diffusion of new products and processes.

The *S*-shaped diffusion curve can be used to estimate the probability that any given firm or individual will be using an innovation at some point in time. The mathematical equation that depicts the curve of Figure 16.5 is given by

$$P(t) = \frac{1}{1 + e^{-(B+at)}}$$

where $P(t)$ is the proportion of adopters at time t and B and a are the parameters that reflect the rate of adoption for a particular innovation. For example, if $B = -5$ and $a = 1$, then $P(t)$ for t ranging from 1 to 9 is

$$P(1) = 0.02$$
$$P(2) = 0.05$$
$$P(3) = 0.12$$
$$P(4) = 0.27$$
$$P(5) = 0.50$$
$$P(6) = 0.73$$

$$P(7) = 0.88$$
$$P(8) = 0.95$$
$$P(9) = 0.98$$

Thus after nine periods, the innovation will have been adapted by 98 percent of the group involved.

The statistical methods are beyond the scope of this book, but techniques are available that allow the coefficients of B and a to be estimated. Once the parameters of the equation have been determined, the proportion of adoption at a future time can be predicted.

Key Concepts

- The Delphi method is frequently used for making forecasts of technological and environmental change.
- Analogy methods predict trends for one innovation based on historical experience with a related innovation.
- The learning curve reflects cost reductions based on the cumulative output produced by a firm. The curve can be estimated using regression techniques.
- Diffusion of many product and process innovations follows an S-shaped curve. This curve can be used to forecast future rates of adoption.

SUMMARY

Technological change can involve new products, improvements or cost reductions for existing products, or better ways of managing. Technological change can be represented as a leftward shift of the production function. Laborsaving change economizes on human inputs and capital-saving technology reduces the need for capital.

Labor productivity is the ratio of output to labor input. Increases in labor productivity may reflect technological change or increases in human capital and the stock of capital goods. A better measure is total factor productivity, which compares changes in output with changes in both labor and capital inputs. Over time, the single most important source of economic growth has been technological change.

Concentrated industries with large firms may facilitate technological change because of the large investments required, scale economies associated with research and development, and the ability of firms with market power to capture the rewards of innovation. The increased dynamic efficiency of firms with market power may offset any technical and allocative inefficiency that exists. However, small firms may provide an environment that is better suited for the development of new ideas. Also, firms facing competition may be forced to innovate to survive. Firms in oligopolistic markets may engage in wasteful duplication of R&D efforts.

Although market structure can affect the rate of technological change, the technology used in an industry can also affect market structure. Changes in technology have made the telecommunications and computer industries much more competitive.

Innovation involves taking new ideas and transforming them into useful products and processes. Diffusion is the process whereby innovations become available throughout society. An innovation should be adopted by a firm if the present value of profits

exceeds the initial cost. For new processes, the keys to successful innovation are secrecy, maintaining some lead time over competitors, and taking advantage of economies of scale and scope. For new products, keys include being able to effectively market the product and providing good service. Innovation is more likely to succeed when the firm remains within its core competencies.

Patents provide a 17-year period during which innovators are provided legal protection from competition. To be patentable, the invention must be new, useful, and nonobvious. U.S. patents are awarded based on who is first to conceive of an idea. A patent may be invalidated by a court challenge.

The Delphi method has often been used to forecast the nature and timing of new technology and future social conditions. Another approach is to forecast one event based on historical experience with a similar or related event.

For some new products, average cost may go down with increases in cumulative output. This learning curve can be estimated by using regression techniques. Diffusion of new ideas may be represented by an *S*-shaped curve. The proportion of individuals or firms that will have adopted the innovation at any time can be predicted by estimating the parameters of the curve.

Discussion Questions

16-1. Could the development of a completely new product be thought of as shifting the production function? Explain.
16-2. Does most technological change tend to be laborsaving or capital saving? Explain.
16-3. What is human capital and how does it enhance labor productivity?
16-4. How could a large firm create a work environment that would allow new ideas to flourish?
16-5. How has technological change affected market structure in the automobile industry?
16-6. Innovators have the advantage of being first. Are there any advantages to being an imitator rather than the first firm to market a product or use a new process? Explain.
16-7. Recently, biologists have learned to manipulate genes. Should such discoveries be patentable? Why or why not?
16-8. Which of the trends listed in the *Megatrends 2000* case study seem accurate today? Are there some that appear to be inaccurate? Explain.
16-9. How does rapid technological change affect the value of input/output analysis as a tool for forecasting?
16-10. Think about the diffusion study for hybrid corn seed. Why did some farmers use it almost immediately, while others waited almost 10 years?

Problems

16-1. A firm's production function is given by $Q = 20K^{0.5}L^{0.5}$. The price of capital is \$10 and the price of labor is \$5.
   a. Determine the expansion path. (Note: The expansion path is discussed on pages 209–211.)
   b. Assume that neutral technological change improves efficiency by 10 percent. Write the new production function and determine the new expansion path.

16-2. A firm's production function is given by $Q = 50K^{0.2}L^{0.9}$.
   a. What do you know about returns to scale?
   b. Assume that neutral technological change improves efficiency by 25 percent. How does this affect the answer to (a)?

16-3. Fawson Enterprises uses labor and materials to produce its product. In 1997, the firm used 10,000 hours of labor and 8,000 pounds of materials to produce 4,000 units of output. In 1998, 9,800 hours of labor and 7,900 pounds of materials were used to produce 4,200 units of output. In both years the price of labor was $20 per hour and materials cost $15 per pound.
   a. What was the percentage increase in labor productivity between 1997 and 1998?
   b. What was the percentage increase in total productivity between 1997 and 1998? Hint: Consider cost per unit.

16-4. A truck is driven 20,000 miles per year. Considering the fuel and maintenance expense, it is estimated that the vehicle costs $0.30 per mile to operate. A new truck would cost $15,000, but operating costs would be only $0.20 per mile. The trade-in value of the old truck is $4,000. The owner uses a planning horizon of 8 years and a discount rate of 12 percent. Should the new truck be purchased? Assume that fuel and maintenance costs are paid at the end of the year.

16-5. A homeowner is considering spending $1,000 to purchase a new fuel-efficient furnace. In a typical year, the cost of heating with the existing furnace is $400. The new furnace will cut heating costs by 25 percent. The owner expects to live in the house for 5 more years and believes that the fuel-efficient furnace will increase the value of the home by $700 when it is sold. Assume that there is no difference in maintenance costs and that a discount rate of 10 percent is used to evaluate the decision. Also assume that all heating bills are paid at the end of the year. Should the new furnace be purchased?

16-6. An office wants to upgrade its computers by replacing the old processors. The new processors cost $1,000 each, and the old processors have a salvage value of $200. Because of their greater speed, each of the new processors is expected to save $600 per year. A 2-year planning horizon and a 10 percent discount rate are used to make the analysis. Assume that the savings accrue at the end of each year. Should the new processors be installed?

16-7. The learning curve is estimated to be $AC = 5.00Q^{-.1}$, where AC is average cost and Q is cumulative output. What is the estimated average cost for 800 units of output? For 900 units?

16-8. The diffusion curve for a new production process is given by $P(t) = 1/(1+e^{-(B+at)})$, where $P(t)$ is the proportion of firms using the process at time t, $B = -3$, and $a = .5$. About how many years will it take until about 20 percent of firms are using the process? Until 50 percent are using it?

Problem Requiring Calculus

16-9. A firm's production function is given by $Q = 30K^{0.7}L^{0.5}$.
   a. Calculate the marginal products of labor and capital.
   b. Assume neutral technological change improves efficiency by 20 percent. Recalculate the marginal products.

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.

16-10. Bailey Manufacturing is contemplating adopting a new quality control system. The incremental revenues and costs over the firm's 12-year planning horizon are as shown here. If the firm uses a 10 percent discount rate for decision making, should the innovation be adopted? Repeat the exercise for discount rates of 5 and 15 percent. What discount rate would make Bailey essentially indifferent about using or not using the quality control system?

Year	*Profit*	*Cost*
0	0	500
1	5	15
2	10	15
3	15	15
4	25	15
5	40	15
6	60	15
7	90	15
8	100	15
9	110	15
10	115	15
11	120	15
12	120	15

16-11. A firm has collected the following data on average costs and cumulative output:

Average Cost	*Cumulative Output*
$150	100,000
140	150,000
132	175,000
127	200,000
124	225,000
122	250,000
122	275,000

a. Estimate the firm's learning curve.
b. What is the difference between the actual and predicted average cost for 200,000 units of cumulative output?
c. What is the predicted average cost for 300,000 units of cumulative output?

CHAPTER 17

Locating the Firm in a Global Economy

PREVIEW

The economic analysis developed thus far neglects an important factor—the effect of the spatial dimension and its implications for locating the firm. Locating a business in the right place is important because the cost of moving output and people across space is significant. For example, the best site for a retail store generally is one that is in close proximity to a large number of people who are the potential customers of the store. In contrast, a manufacturing firm may combine raw materials from several sources and ship manufactured output to customers at other sites. In this case, an important location criterion is the cost of shipping raw materials relative to the cost of shipping final output.

Another consideration is that all firms employ workers who must travel from their homes to the firm each working day. The firm must locate in close proximity to that labor supply or face the prospect of paying premium wages to compensate workers for traveling long distances and/or providing housing and other amenities for workers at the employment site. The latter are characteristic of installations located in remote locations, such as offshore oil drilling platforms and some mining operations.

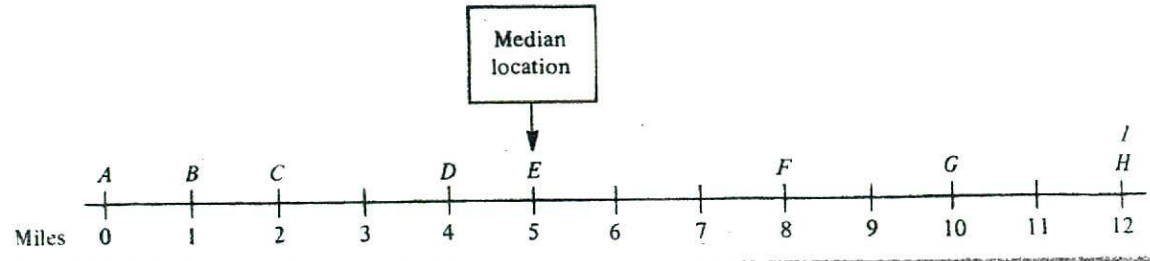
Clearly, some locations for the firm are better than others, and there are numerous examples of business failures that can be directly attributed to the selection of a poor location. In this chapter, five topics in location theory are discussed. First, the basic theoretical principles of industrial location are considered. Next is a discussion of the determination of the market area for a firm; that is, given the location of firms in a region, what share of the market will accrue to each? The third section covers the principle of threshold analysis, which explains why certain economic activities are found in some areas but not in others. In the next section, the relative importance of a variety of location factors is discussed. Finally, principles guiding the location of the firm in the global economy are developed.

BASIC LOCATION PRINCIPLES

In this section, the fundamental principles of industrial location are outlined.[1] Consider, for example, the problems that would be associated with locating a manufacturing firm that used several different raw materials, each of which could be obtained from suppliers in many parts of the country, and that sold several different products to many customers in a number of locations. Seeking the location that minimized the total transportation costs for raw material and output could become a very complicated problem.

The simple models discussed here, however, provide the flavor of location theory and illustrate some important location principles. One principle is that there is a tendency for firms and individuals to locate together in particular areas. Even the most casual observer of geography is struck by the concentration of economic and human activity in cities. There are a variety of reasons for this concentration, but one of the most important is economic in nature. By locating in close proximity, the costs associated with moving people, goods, and information are reduced. Furthermore, these concentration forces tend to be mutually reinforcing and can cause a cumulative buildup of population and economic activity in an area. For example, suppose that a shopping

[1]For an excellent summary of the basic principles of location theory, see W. Alonso, "Location Theory," in *Regional Policy: Theory and Applications*, ed. W. Alonso and J. Friedman (Cambridge, Mass.: MIT Press, 1975).



center locates near a concentration of people. As a result of the employment and shopping opportunities provided by this center, more residents are attracted to the area, which creates demand for even more stores. This cumulative process is one explanation for the development of urban areas.

In the following discussion, several alternative models are developed to demonstrate the optimal location for a firm under specified market conditions. In all these models, there is a tendency for the firm to locate at a central point, such as an urban area where output is sold, or at the site of a supply of raw materials. Although the models are very basic, they illustrate many of the key principles of location theory.

Locating in a Linear Market

If demand for a firm's output does not vary with location, the problem of locating a plant or service center reduces to one of cost minimization. Suppose that the letters *A* through *I* in Figure 17.1 represent households located on a highway. Such a distribution of customers is referred to as a *linear market*. Assume that each customer must be served once each month by delivering one truckload of output (e.g., coal or fuel oil for heating) to each home. Where should a firm locate its distribution center to minimize the total transportation costs of servicing these consumers?

The cost-minimizing solution is to locate at the median point, where there are as many customers on either side of the distribution center. Costs are minimized at this point because moving the firm in either direction adds more distance to people on one side than it subtracts from people on the other. The median location is at the 5-mile mark, where there are four customers on either side of the firm. The total mileage required to serve all customers is 70 miles, computed as the sum of a 10-mile round trip to *A* plus an 8-mile round trip to *B*, and so on. No other location will allow each customer to be served and result in fewer than 70 miles driven.[2]

[2]A seemingly logical, but incorrect, answer would be to build the facility at the average or mean location. To find the mean location, start at point zero on the highway and find the average distance traveled if all customers are served. That is, customer *A* is zero miles away from the endpoint, *B* is 1 mile away, and so on. The total number of miles is $0 + 1 + 2 + 4 + \cdots + 12 = 54$, and thus the average distance is 6 miles (i.e., 54 miles divided by nine customers). Using this criterion, the firm would locate at the 6-mile mark and would make nine trips each period, totaling 72 miles. That is, one 12-mile round trip would be made to *A*, one 10-mile round trip to *B*, and so on. But this solution requires 2 additional miles of driving compared to the median solution.

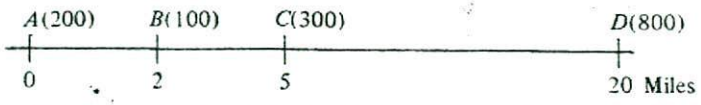


FIGURE 17.2 Number of Consumers at Each of Four Cities Along a Highway

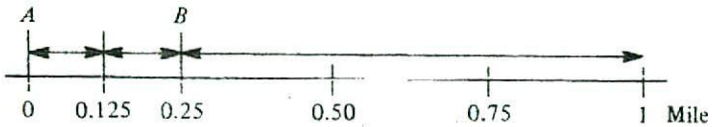


FIGURE 17.3 Initial Location of Ice Cream Vendors

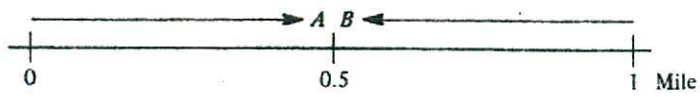
Consider another example of a linear market. Suppose that a firm is seeking a location to serve four cities, *A, B, C,* and *D,* along a highway. The number of customers in each city is shown in Figure 17.2. Using the principle of median location, the firm will minimize transportation costs by locating at *D.* Technically, the median location would be at some point on the west side of city *D,* where there are as many customers to the west as there are to the east.

Now, extend this analysis to another case where two firms are seeking locations along a linear market. Suppose that this market is one mile long with customers distributed uniformly along it. An example would be swimmers at a beach. Two vendors, *A* and *B,* sell ice cream to these swimmers using easily moved stands. Each consumer buys one ice cream bar each day, the price is the same at both stands, and consumers will patronize the closest vendor. Assume that the sellers initially set up their stands at the 0- and 0.25-mile marks along the beach, as shown in Figure 17.3.[3]

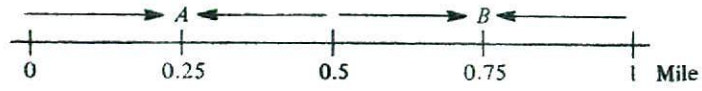
Because both sellers charge the same price and the swimmers go to the closest stand, initially *B* will have 87.5 percent of the market because all swimmers to the right of the 0.125-mile mark will buy from *B.* But this is not an equilibrium location pattern. As the stands are easily moved, *A* can capture most of *B*'s market by moving his stand just to the right of *B,* say, to the 0.26-mile mark. This would give *A* about 74 percent of the market. It is likely *B* would respond by moving to the right of *A.* Then *B* would move just to the right of *A,* capturing perhaps 73 percent of the market. Such continuous relocation would continue until both stands were located adjacent to one another at the 0.5-mile mark, as shown in Figure 17.4a. This is the equilibrium location pattern in this market because both sellers now have 50 percent of the market and neither seller can increase market share by moving.

Although this example is somewhat simplified, one can think of many cases where a number of competing firms all locate in one part of an area. For example, some urban areas have a theater district where a number of stage and movie theaters are located together. Shopping centers cluster a number of retail stores together. Once again, principles of location economics suggest a concentration of economic activity.

[3]This example was originally developed in a classic article by H. Hotelling, "Stability in Competition," *Economic Journal* 39 (1929): 41–57.



(a) Free market equilibrium location pattern



(b) Minimum transportation cost location pattern

FIGURE 17.4 Alternative Locations for Ice Cream Vendors

Although the location of both vendors at the 0.5-mile mark is the free-market equilibrium solution, it is not socially optimal because total transportation costs for all swimmers walking to the ice cream stand are not minimized. With the swimmers evenly distributed along the beach, the average distance to the nearest ice cream stand will be 0.25 mile if both stands locate at the 0.5-mile mark. This distance can be reduced by one-half, to 0.125 mile, by a regulation requiring that the vendors locate at the 0.25- and 0.75-mile marks, as shown in Figure 17.4b. In this case, the sellers would be located at the median locations of each half of the market. Note that both sellers still retain 50 percent of the market, but the average distance traveled by swimmers is reduced. Generally, free-market solutions to economic problems result in socially optimal outcomes. This is one case where that principle does not hold.

Thus far it has been assumed that the demand for ice cream by each swimmer is one per day and does not depend on the location of the stands. This is somewhat unrealistic because the quantity demanded would probably be inversely related to distance from a stand. A swimmer located near the stand might have three or four bars, whereas a swimmer located 0.5 miles away might decide that it is simply too far to walk, and thus not purchase any. That is, the price, including transportation costs, is too high for that individual.

Assume that the maximum distance a buyer would walk is 0.25 miles. In this case, the vendors would each sell twice as many bars by locating at the 0.25- and 0.75-mile marks (each having one-half the market) than by both locating at the center. In the latter case, each would have 50 percent of the entire market (i.e., one would have that part from the 0.00- to the 0.5-mile marks, while the other one would have that part of the market from the 0.5- to the 1.00-mile marks). Given these demand conditions, which probably are more realistic than those assumed in the previous example, the profit-maximizing and the socially optimal solutions are the same.

In general, there is a tendency for firms to locate in the middle of market areas. In larger urban areas, there typically is a central business district that serves the entire city with certain goods and services and a number of smaller business districts (e.g., shopping centers) that serve a submarket of the city. In any case, the principle of median location helps to explain the concentration of these businesses at particular points.

The principle of median location is one reason why the urban centers in the world have grown so much in the past 50 years. These cities are the median, or at least central, locations for many types of economic activity. For example, the downtown areas and

suburban shopping centers of urban areas often have a number of the same type of store located very close together. All have sought the median location at the center of a market area. Obviously, one finds stores scattered at various points in urban and even rural areas. However, the growth of urban centers and the concentration of economic activity at points within those centers suggests a strong tendency to locate businesses at or near the median location within those markets.

Case Study

Location Theory, Product Attributes, and the Personal Computer Industry

Firms producing similar goods or services tend to locate in close spatial proximity. Similarly, in many industries, there is little variation in the attributes of the products sold by the firms. That is, there is close proximity of attributes. The personal computer industry is a good example. When the industry began to develop in the early 1980s, there were several major manufacturers producing machines that were not compatible. For example, Osborne, Kaypro, Victor, Apple, and IBM all built personal computers, but each required some specially designed peripheral equipment and software for one generally could not be used on the others. However, IBM quickly became the industry leader, and attributes of the IBM-PC became the standard against which other computers were built. As a result, other computer manufacturers began to develop and market their computers in terms of their "IBM compatibility." This compatibility stressed the extent that these computers would run software written for the IBM, but also involved features such as the ability to use expansion boards and peripheral devices designed for the IBM machine. Once basic compatibility was established, other features, such as lower price or greater memory, could be promoted in order to differentiate the product from the IBM-PC.

The task of these competing manufacturers was a difficult one. They had to position their product close enough to the IBM-PC to convince potential purchasers that it could do virtually anything that the IBM computer could do. But the product had to be differentiated enough to establish a reason for buying it instead of the IBM product. Advertisements often stressed greater speed or lower price than comparable IBM products.

At the present time, there are many manufacturers of standard desktop personal computers that are all "IBM-compatible." In addition, a number of other manufacturers have given up trying to compete in this market. As suggested in chapters 9 and 10, the economic profits being earned by Apple, IBM, and some of the other early entrants attracted numerous firms into the industry, including AT&T for a while. Although the market has grown tremendously, fierce competition has resulted in much lower prices, and most firms in the industry probably are not now earning economic profits.

The tendency for product attributes of different firms to be similar is not unique to the computer industry. Indeed, it is a characteristic of most mass-produced products. A television set must have most of the features of other makes. New automobile models

seldom represent radical departures from competing models that are already on the market. Most new textbooks try to retain most of the features of texts that have been successful in the past. In each case, the explanation is the same as for the locational clustering of firms. If the product is too different, that is, too far from the "center" of the market, it is likely to attract fewer customers than if it has more traditional attributes. ■

Key Concepts

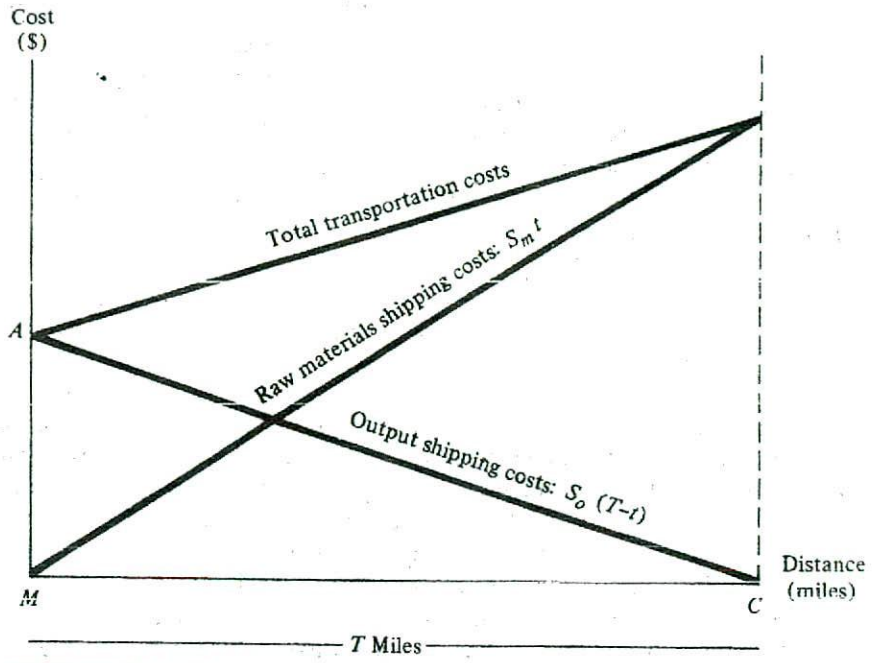
- Location of business is important because the cost of moving resources and output across space is significant.
- In general, principles of location theory suggest that firms and individuals benefit by locating close together and that these location forces tend to result in concentration of economic activity in urban areas.
- If demand does not vary with location, two firms serving a linear market will both tend to locate at the center of the market, resulting in higher transportation costs for their customers than if they located at separate points.
- If demand is inversely related to distance from the seller, the two firms will tend to locate in the center of each half of the market.

Firm Location: One Market and One Raw Materials Source

In the previous examples, the focus was on minimizing only the transportation costs associated with sending final output to consumers. For example, no consideration was given to the cost of shipping ice cream bars to the vendors. This section considers a more realistic but still simple case of location decision for a firm that obtains raw materials at one site (M), processes them, and distributes to customers in a city (C). As shown in Figure 17.5, the raw materials and market sites, M and C, respectively, are T miles apart.

Assume that production cost, the price of output, and the quantity sold are the same regardless of where the firm locates the plant. The only variables are the total costs of transporting raw materials and output. Therefore, the problem reduces to determining the location that will result in minimum total transportation costs.

Let S_m be the cost per mile of shipping enough raw materials to make one unit of output (referred to subsequently as one unit of raw material) and S_o be the cost per mile of shipping one unit of output. The cost function for shipping raw materials is $S_m t$, which determines the cost of shipping one unit of raw material from M to any location t miles to the right of M. If the plant is located at the raw materials site, then $t = 0$ and the cost of shipping raw materials is zero. The output shipping cost function is given by $S_o(T - t)$, which defines the cost of shipping one unit of final product from any location $(T - t)$ miles to the left of C to that city. If the plant is located at C, the value of $(T - t)$ is zero and the shipping costs for output are zero.



At any intermediate point between M and C, there are shipping costs for both raw materials and output. The sum of these two costs is defined as total transportation costs. As price and other production costs are assumed to be constant at all locations,[4] the problem reduces to choosing that site with the lowest total transportation costs. In the example shown in Figure 17.5, the per-mile cost of shipping one unit of raw materials is assumed to be greater than the cost of shipping one unit of output. Thus, total transport costs are minimized at the raw materials site M. In contrast, had transportation cost per mile for raw material been lower than for one unit of output, the lower total transportation costs would be achieved by locating at C.

This principle can be demonstrated mathematically. The total transport cost (TC) of locating the plant at any site t miles to the right of M is the sum of shipping costs for raw material and output, that is,

$$TC = S_m t + S_o(T - t)$$

or

$$TC = (S_m - S_o)t + S_o T$$

Recall that the plant will be located at some point, t, where $0 \leq t \leq T$. As the objective of the plant location decision is to minimize TC, it should be clear that if $S_m > S_o$, the value of t should be made as small as possible. This is accomplished by locating the plant

[4]Clearly, production costs and demand may vary among alternative locations, but this assumption simplifies the problem, allowing the analysis to focus on the transportation-related issues.

at M, where $t = 0$, as shown in Figure 17.5. Alternatively, if $S_m < S_o$, total costs are minimized by making t as large as possible (i.e., $t = T$) and locating the plant at C.

If $S_m = S_o$, the total cost function would be horizontal and the firm would be indifferent about locating at any point between M and C. However, if there are costs associated with the loading and unloading of raw materials and/or output onto trucks, rail cars, or ships, then even if $S_m = S_o$, any intermediate location will have higher costs than at M or C because of the additional terminal costs incurred. For example, by locating at M, the loading of raw materials is avoided, and by locating at C, the loading of final output is avoided. If an intermediate location is selected, both of these costs will be incurred. Thus, industrial location tends to take place either at a raw materials site or at the market. This phenomenon once again suggests the natural tendency for economic activities to be concentrated at certain places.

In general, products that tend to be "weight losing" in the production processes are associated with locations at the raw materials site. For example, the processing of gravel involves screening and washing large quantities of rock, dirt, and other debris to separate that part of the load that qualifies as gravel. The usable gravel may only be one-third of the material processed. Invariably, this process takes place at the raw materials site, and then the finished product is taken to the marketplace, often to the site of a construction project. In contrast, "weight-gaining" activities tend to locate at the marketplace. For example, soft-drink bottling plants add small quantities of concentrate to large volumes of water to manufacture soft drinks. As water is available in virtually all locations at a relatively low cost but is expensive to transport, the economics of location dictate that bottling plants be located in market areas. Indeed, bottling plants are found in virtually every area of the country. The market area of each plant is relatively small because it is too expensive to ship bottled soft drinks very far, since they are more than 90 percent water.

Case Study

Locating the Steel Industry

In the early part of the twentieth century, location decisions in the iron and steel industry were based primarily on minimizing the costs of the raw materials used in production, specifically coal and iron ore. At that time, it took approximately 4 tons of coal and 2 tons of iron ore to manufacture 1 ton of steel. Clearly, this was a weight-losing process, and the shipping costs of raw materials dominated the location decision. The steel industry also uses large quantities of water—about 65,000 gallons per ton of steel produced—for cooling and processing. Although much of this water is continually reused, there must be large supplies available. Thus, locations on rivers or lakes were definitely preferred. Because western Pennsylvania had abundant supplies of coal and water available at low prices, virtually the entire U.S. steel industry located there, and Pittsburgh became the nation's steel capital.

Improvements in technology steadily reduced the amount of coal required to process a ton of iron ore, and by the mid-1970s less than 1 ton of coal was used per ton

of ore. While the industry remained tied to raw materials sources, it became relatively more important to locate near the sources of iron ore, much of which was mined in the Upper Great Lakes area.

Taking advantage of low-cost water transportation for both coal and iron ore and large water supplies, major steel-producing complexes developed at Cleveland, Youngstown, Detroit, Buffalo, and Chicago—all cities located on the Great Lakes. Smaller complexes developed later at St. Louis, Birmingham, and Fontana, California.

More recently, the use of scrap metal as a raw material has become important. The existence of large supplies of scrap metal in Detroit and Chicago helped the steel industry to grow in those areas. In fact, relatively small steel mills that depend almost entirely on scrap metal have been developed in places never thought to have potential as steel industry locations. ■

Key Concepts

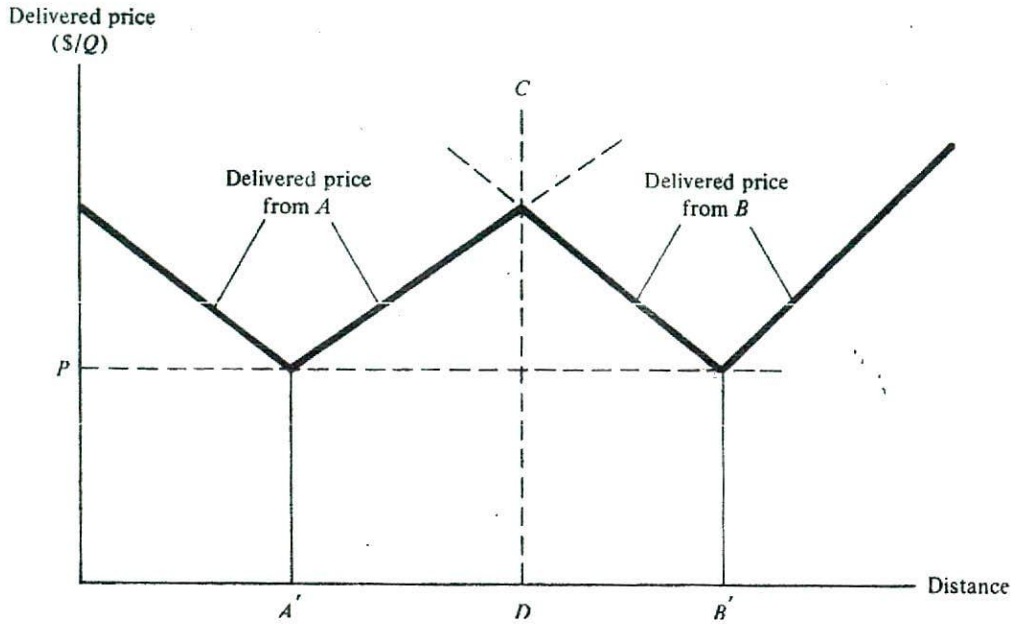
- If other factors are held constant, a firm that uses raw materials from a site *A* and sells output in the market at *B* will locate at *A* if the cost of shipping raw materials is more than the cost of shipping final output. Otherwise, the firm will locate at the market, *B*.
- Even if the transportation costs are the same for raw materials and finished output, the lowest-cost location will be at either the raw materials site or the market because additional loading and unloading costs would be incurred at any intermediate location.
- Production activities that are weight losing tend to locate at the source of raw materials, whereas production activities that are weight gaining tend to locate at the market.

MARKET AREA DETERMINATION

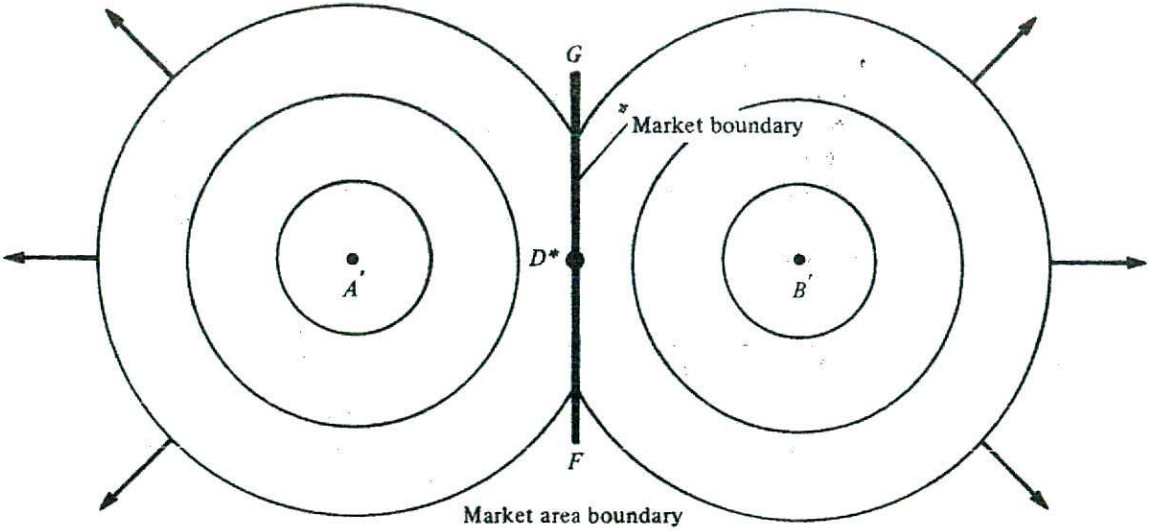
The market area for any one seller will depend on relative production and transportation costs. In the following discussion, market areas for two competing firms will be determined under different assumptions relative to those costs.

Market Area: Equal Production and Transportation Costs

Suppose that two competing firms, *A* and *B*, have located production facilities in a region and that both have the same production and transportation costs. The price at the plant is set equal to production costs, and transport costs are paid by the consumer. Consumers will buy from that plant for which the delivered price (i.e., price at plant plus transport cost) is lower. Figure 17.6a shows delivered price functions for firms *A* and *B* located at



(a) Delivered price schedules



(b) Market areas corresponding to delivered price schedules in figure 17.6a

points A' and B' along the distance or horizontal axis. Production cost is $0P$ at both plants. Any buyer located at point A' or B' pays $0P$, but more distant buyers must pay $0P$ plus the transportation cost from the plant. For example, if S is the per-mile transportation cost per unit of output, a buyer located t miles from A must pay $0P + St$. Thus, the delivered price functions shown in Figure 17.6a have a slope equal to S.

Because consumers seek the lowest price, all those located to the left of point D will buy from A and those to the right of D will buy from B. Looking down on this two-dimensional market area, as shown in Figure 17.6b, it is seen that the line FG separates the two market areas. Along this market boundary, the delivered price is the same for both plants. To the right of this line, the price is lower for B, and to the left, it is lower for A. Point D in Figure 17.6a corresponds to point D^* in Figure 17.6b.

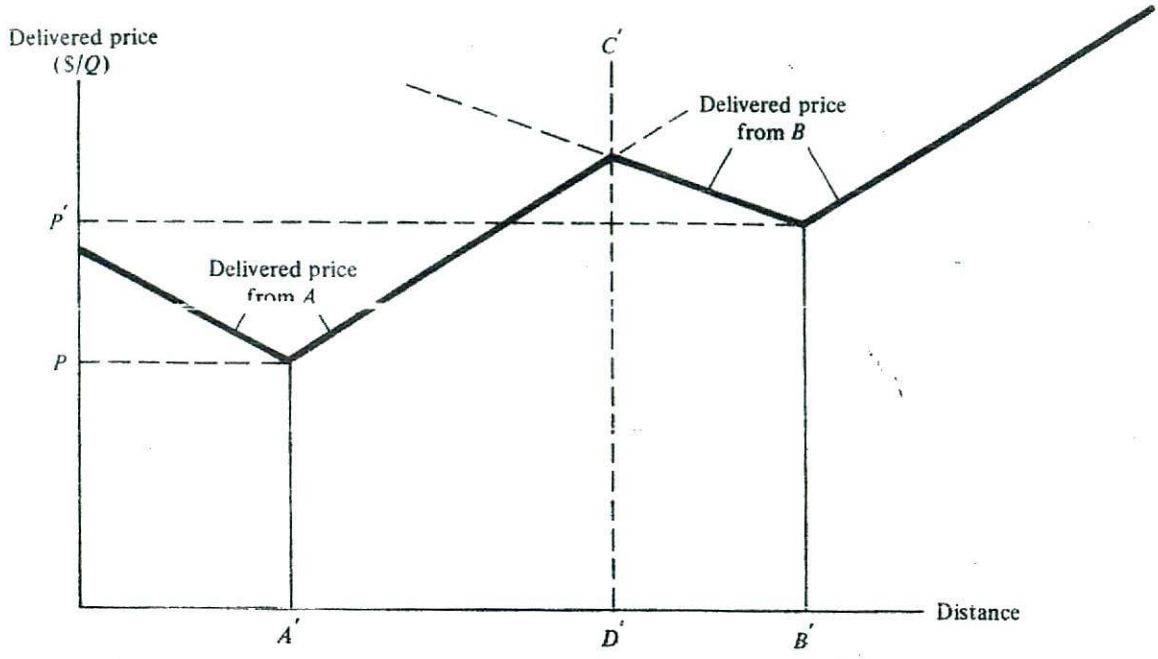
Market Area: Unequal Production Costs—Equal Transportation Costs

Suppose that firm B's production cost per unit increases, while the costs at firm A remain the same. Figure 17.7 depicts the market areas for the two firms where A is able to produce at a lower cost than is B. As a result, the price at the plant is lower for A than for B. Per-mile transport costs are assumed to remain equal for both firms. Thus, the slope of the delivered price functions is the same, but the cost function for firm B is higher by the amount of the production cost differential between the two firms. This change results in A's market area increasing significantly. That is, now some customers who are geographically closer to B will buy from A because the lower production cost has more than offset the greater transportation cost. Also, the boundary between the markets, the curve $F'G'$ in Figure 17.7b, is now nonlinear. The point D' in Figure 17.7a corresponds to point D^* in Figure 17.7b.

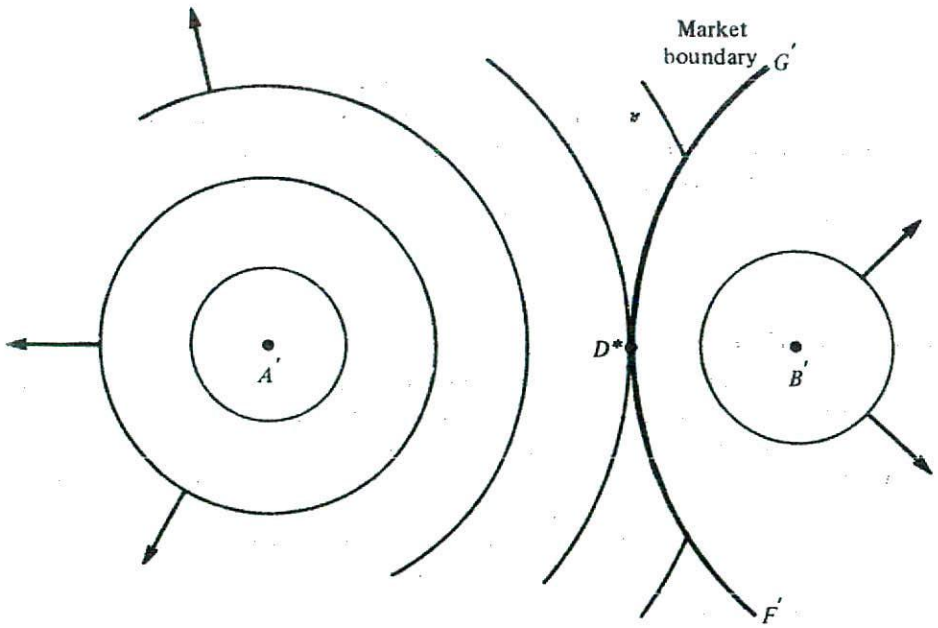
The large share of the U.S. automobile market taken from American producers by Japanese producers located thousands of miles away is an example of this situation. Japanese automakers had lower production costs that more than enabled them to offset the additional transportation costs of shipping their output to American buyers. Thus, they took over a significant share of the market for cars in the United States.

Market Area: Unequal Production and Transportation Costs

Next, suppose that both production and transportation costs are lower for A than for B. Transportation costs may be lower for A because it has developed access to lower-cost barge and/or rail service, whereas B may be restricted to using higher-cost truck service. Alternatively, B may simply have higher costs due to obsolete equipment and/or poor transportation management. The determination of the market area for each firm is depicted in Figure 17.8. Note in Figure 17.8a that B's market area is restricted to the area between D'' and E'. Even though customers to the right of E' are closer to B' than to A', the delivered price from A actually is lower. A's lower production and transport costs have reduced B's market area to a small circular area. The two-dimensional perspective of this market area analysis is shown in Figure 17.8b. The market area for firm B is the small circular area B^*B^*. Firm A captures the remainder of the market areas.

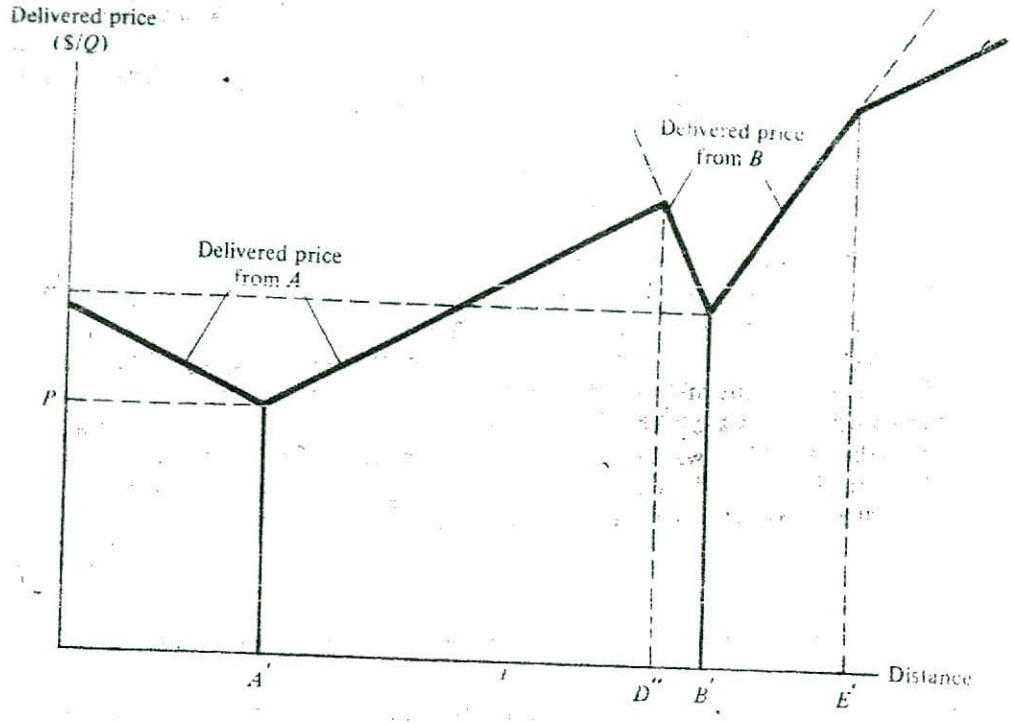


(a) Delivered price schedules

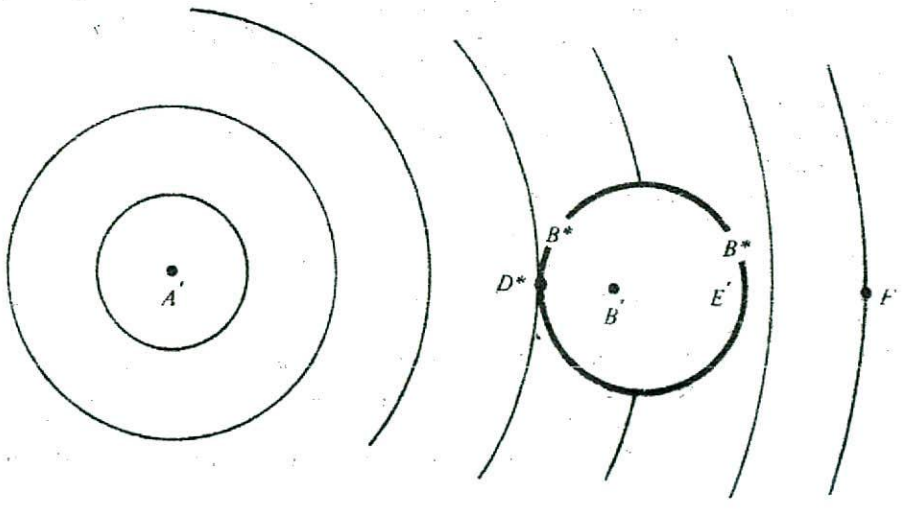


(b) Market areas corresponding to the delivered price schedules in figure 17.7a

FIGURE 17.7 Price Schedules and Market Areas: Firms with Unequal Production Costs but Equal Transportation Costs



(a) Delivered price schedules



(b) Market areas corresponding to the delivered price schedules in figure 17.8a

FIGURE 17.8 Price Schedules and Market Areas: Firms with Unequal Production and Transportation Costs

It is possible that further reduction in A's production and/or transportation costs could occur to the point that the delivered price function for firm A would intersect that for firm B below the level of B's production cost $0P'$. In that event, firm A would capture the entire market area and B would be out of business.

Key Concepts

- The market boundary for two firms is a function of relative production and transportation costs and is defined by that set of points where the delivered price is the same for each firm.
- If production and transport costs are equal for both firms, the market boundary will be a straight line midway between the two producers.
- If production costs are higher for one firm but transportation costs are equal, the market boundary will be a curved line around the higher-cost plant.
- If both production and transportation costs are higher for one plant, its market area will be reduced to a circular area around its plant site.

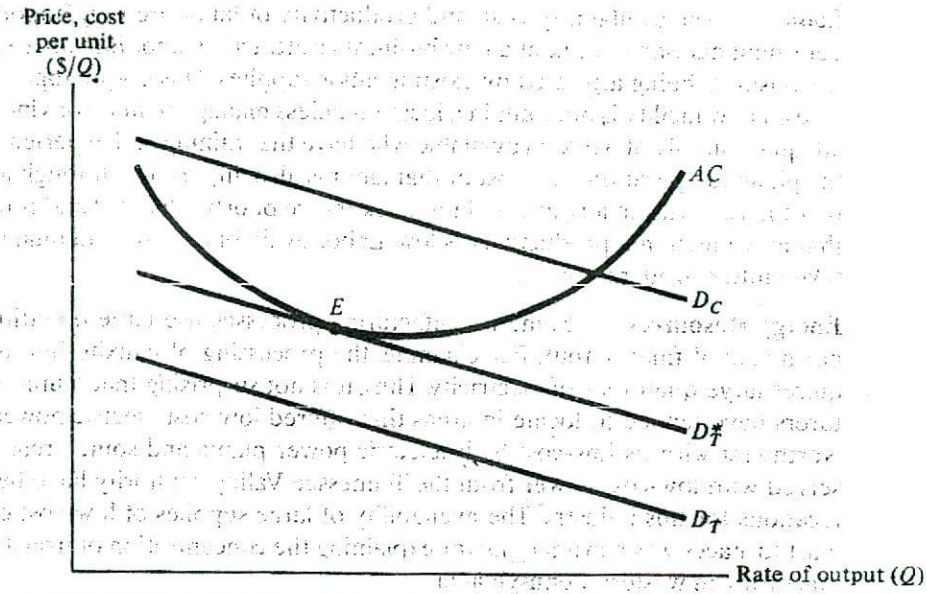
THRESHOLD ANALYSIS

There is a consistent hierarchical arrangement of businesses among cities of various sizes. For example, automobile service stations and convenience stores are generally found in even the smallest communities. Conversely, to find symphony orchestras and gourmet restaurants, one must usually go to large cities. The reason for this is a matter of economics. The demand for symphony orchestras in small cities simply is not great enough for revenue that would cover costs. When the market has become large enough so that revenue is at least as great as cost, it is said that the threshold level for the activity has been reached.

Figure 17.9 shows hypothetical demand curves for a gourmet restaurant in a small town (D_T) and in a city (D_C). The average cost curve for the restaurant also is shown. It is assumed that the same cost conditions would prevail in both places. Note that average cost is above the small-town demand curve at all points. Thus, it is not possible for the restaurant to break even under these conditions. In contrast, part of the demand curve for the city is above the cost function and therefore there are output rates for which the restaurant is profitable.

If the small town increased in size, demand for restaurant services would probably increase. If demand increased to D_T' which is tangent to the average cost curve at point E, a threshold has been reached, and it would be possible for the restaurant to break even. Of course, this threshold also could have been attained by a downward shift in the average cost curve.

As population centers increase in size, not only do more businesses of the kind already there develop (i.e., another service station or drive-in restaurant), but new kinds of businesses appear. For example, banks or small department stores are often found in small cities but not in the smallest towns. As the latter grow, these activities appear when the threshold level is reached.



SELECTING AN INDUSTRIAL LOCATION

Obviously, the process used by managers to select a location for a store or plant is more involved than that suggested by the preceding theoretical discussion, but the process is consistent with that theory. Typically, the firm has a general idea about the set of possible locations for the facility. For example, a firm may find that a particular regional market such as the West Coast has grown sufficiently that a production facility can now be justified in that area. Often, a professional consultant with a particular expertise in facility location is brought in to make the evaluation. The firm provides that consultant with a rating of the importance of all location factors. For example, there may be specific requirements for rail service, large quantities of skilled labor or other input, or educational facilities that are found in only a few places.

In general, all significant factors that will influence the profitability of the proposed facility are evaluated at each location under consideration. Because both revenue and cost may differ at each location, the analysis must consider each location's attributes as they affect these two key variables. These locational factors are classified as primary and secondary.

Primary Location Factors

The locational attributes described here are fundamental in the decision to locate an industrial facility.[5] Although for particular firms some are more important than others, a significant shortfall in an area's ability to provide even one of these may greatly reduce the attractiveness of that site.

[5] For a detailed discussion of the industry location process and the importance of each locational attribute, see E. W. Miller, *Manufacturing: A Study of Industrial Location* (University Park, PA: Pennsylvania State University Press, 1977).

Labor The availability, cost, and productivity of labor are very important location determinants. Some think of a new business as attracting labor into an area rather than the business being attracted by existing labor supplies. This does happen, but in many cases, a new facility is not built in a location unless management is convinced there is an adequate supply of workers available who have the training and experience needed for the planned operation. The cost of that labor is also important, although low labor cost is not necessarily an advantage if the workers are poorly educated and trained, because that may mean that productivity is low. Labor availability, cost, and quality must all be taken into consideration.

Energy Resources Some manufacturing processes use large quantities of energy per dollar of final output. For example, the processing of bauxite into aluminum requires large quantities of electricity. Thus, it is not surprising that aluminum manufacturers have tended to locate in areas that offered low-cost electric power. The Pacific Northwest with its low-cost hydroelectric power plants and some areas in the South served with low-cost power from the Tennessee Valley Authority have been attractive locations for this industry. The availability of large supplies of low-cost coal for firing blast furnaces was a primary factor explaining the concentration of iron and steel manufacturing in western Pennsylvania.

Transportation Transportation cost also is an important factor in industrial locations decisions. In the early history of the United States, locations near water and rail transportation were very important. This is evidenced by the rapid rate of growth of cities that offered good access to these transport modes. The relative growth of the service sector, which is not heavily dependent on transportation, and the emergence of truck transport have greatly expanded the range of good industrial locations. For example, the manufacturing activity that has developed in smaller communities and in suburban areas of large cities has depended largely on truck transport and the development of the interstate highway system.

Proximity to Markets for Output In general, it is advantageous for firms to be able to serve their customers quickly and at low cost. Clearly, one way to do this is to locate close to these customers. Also, as indicated previously, firms need to locate near their employees, and people need businesses as places of employment. By locating together, the cost of transporting both people to work and products to consumers is reduced, as is the cost of communications. These forces combine to ensure that many firms will locate in population centers. Except for firms that are tied to raw material sites (e.g., where the production process results in large weight losses during manufacturing), there are strong economic forces pushing firms toward locations in urban centers.

Government Regulation Federal, state, and local governments are assuming a more aggressive role in determining where industry can locate. Many local governments have used zoning laws to regulate the location of businesses within the city. More recently, air and water quality rules imposed by federal and state governments have been important in determining where polluting industries can locate. Although such controls have benefits, if taken too far, they discourage new business from locating in an area

and may actually encourage some firms to move. The combination of their tax structure and government regulation of business has resulted in some areas of the country being regarded as probusiness and others as being antibusiness.

Raw Materials Availability Some businesses depend on materials of various types, such as unprocessed raw materials, for use in manufacturing and finished goods for inventory in wholesale and retail establishments. The availability and cost, including transportation costs, of these materials are important location factors.

Secondary Location Factors

There are other factors that influence location decisions but are of secondary importance. Two examples are the physical environment and government attitudes and policies.

Physical Environment The climate, scenery, and environmental quality of an area may affect a location decision, especially if the other more fundamental characteristics are approximately the same as in the other locations being considered. Although there have been reports of an industrial location being selected simply because the company president enjoyed the local scenery, such cases probably are rare. The competitive pressures of the marketplace would drive out firms whose important decisions were based on whim or personal preference rather than economic considerations.

State and Local Government Attitudes and Policies Much has been written about the climate for business in the various states. Some states in the Northeast and Midwest are thought by some to be antibusiness, whereas others, especially in the South, are so probusiness that they offer subsidies to new firms to locate plants in their states. These subsidies may take the form of property tax reduction, free sites for factory construction, low-interest loans, and even the construction and lease of buildings at below-market rates. Despite claims about the importance of these subsidies, most evidence suggests that they play a secondary role in the location decision. Sometimes, however, businesses are able to obtain such subsidies by threatening to go elsewhere even though they have already decided to locate in the area. Their objective is simply to achieve a further reduction of their costs. In other cases, a firm may have decided to locate in a region and then will go shopping among the communities in the region for the best set of subsidies.

It should be recognized that low taxes in an area are not an advantage to a business if they are associated with a low level of public services. Poor schools, inadequate water and sewer systems, and/or limited police and fire protection may be the product of low taxes. These are not the characteristics of a good industrial location, and the firm may have to incur additional costs to offset these inadequate public services. For example, elementary and secondary schools are important to the business because they influence the quality of the local labor force and are an important consideration for managers and other workers who might have to be recruited by the firm. If the schools are not good, it may be very difficult to attract workers to the area. Thus, wage rates may have to be increased substantially to attract these workers, at least partially offsetting the advantage of low taxes.

Case Study

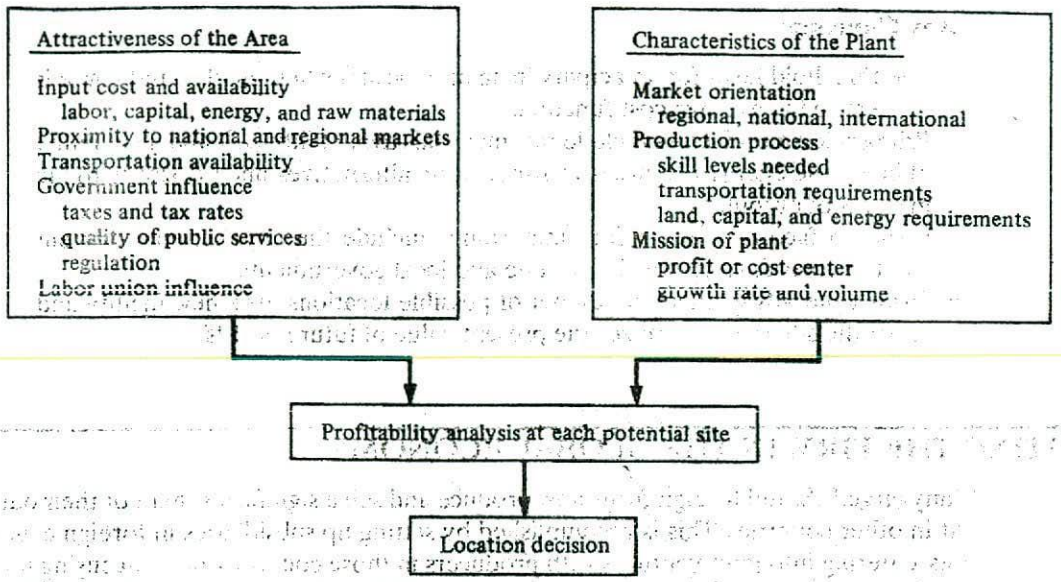
Ranking Industrial Location Determinants

A number of surveys of managers responsible for industrial location decisions have attempted to determine the importance of various factors in choosing an industrial location. The relative importance of these factors as reported in a recent study is shown in the following table.

Geographic location clearly was the most important factor. This attribute captures the effects of being close to markets, to supplies of labor and raw materials, and to sources of specialized business services such as those provided by law and engineering firms. Worker productivity, availability of skilled labor, and a low union profile are also thought to be important. It is interesting that land (i.e., truck) transportation availability is generally regarded as much more important than is rail and water transport. In the past, rail and water would have been regarded as much more important. This change reflects the combination of better highway systems and changing technology in the transportation industry. ■

Location Factor	*Percentage of Executives Surveyed Who Rated the Factor as Being Vitally Important in the Industrial Location Decision*
Geographic location	64
High worker productivity	59
Availability of land transportation	54
Low union profile	49
Stable state government	38
Skilled labor availability	32
Energy sources	30
Raw materials availability	28
Tax exemptions	27
Tax credits	26
Unskilled labor availability	22
Availability of air transportation	21
Water supply	17
Availability of rail transportation	16
Availability of worker training programs	10
Availability of water transportation	5

SOURCE: M. L. Goldstein, "Choosing the Right Site," *Industry Week* (April 16, 1985).



The Industrial Location Decision

Deciding where to locate an industrial facility generally is based on a comparison of the characteristics of that facility to the attributes of the various locations being considered. The first step usually involves an inventory of an array of locational factors at each area. This is followed by an evaluation of how the characteristics of the proposed facility mesh with the area's locational attributes. The objective is to determine the present value of all future profits over the life of the plant at each location under consideration. This process is described in Figure 17.10.

Many government units in the United States have local development groups that promote the industrial development attributes of their area. This is done by advertising in business publications, direct contact and solicitation of industry, and making improvements in the area's locational attributes. Examples of the latter might include improvement of highways, establishment of vocational training centers, and development of land for new plant sites. The promotional activities, sometimes referred to as smokestack chasing, are perhaps the most visible part of the industrial location process.

Firms also spend a considerable amount of time and money evaluating various parts of the country for plant locations. Analysts may develop projected income statements for a proposed plant for each of a number of possible sites. Generally, the site that maximizes the present value of all future profits over the estimated life of the plant is selected.

Key Concepts

- The threshold level for an activity in an area occurs when the demand curve is tangent to the average cost function.
- Primary location factors include raw materials availability, cost and availability of labor and energy resources, transportation alternatives, and proximity to the markets of output.
- Location factors of secondary importance include the physical environment and the attitudes and policies of state and local governments.
- Firms generally evaluate a number of possible locations for a new facility and select the site that maximizes the present value of future profits.

LOCATING THE FIRM IN THE GLOBAL ECONOMY

Many large U.S. and foreign firms now produce and sell a significant share of their output in other countries. This is accomplished by setting up subsidiaries in foreign countries, entering into joint ventures with producers in those countries, or by licensing foreign firms to produce and market their products. Table 17.1 shows the foreign sales of selected U.S. firms. Not only are these sales a significant share of total sales, for many firms their foreign operations have grown more rapidly than their domestic revenues, as international market penetration has increased. The explanation for the expansion of firms throughout the global economy lies in three areas: developing and maintaining raw materials supplies, extending market power, and comparative advantage.

Raw Materials Supplies

Firms that are heavily dependent on particular raw materials need to ensure that they have continued access to dependable supplies. In some cases, materials are only available in foreign countries. An example is bauxite, which is the basic raw material used in producing aluminum. Virtually all of the deposits of this mineral are located outside the United States.

TABLE 17.1 Total Revenue and International Sales of Selected U.S. Corporations, 1990

Firm	*Total Revenue (billions)*	*Foreign Sales (billions)*	*Foreign Sales as a Percentage of Total Revenue*	*Type of Business*
Caterpillar	$ 16.5	$ 8.4	51%	Earthmoving equipment
Ford	147.0	60.3	41	Automotive equipment
General Electric	46.1	19.4	42	Appliances; technology; communications
General Motors	161.4	37.7	23	Automotive equipment
IBM	75.9	36.4	48	Information-processing equipment
Rohm and Haas	3.9	1.8	47	Specialty chemicals
Texas Instruments	9.9	5.4	55	Electronic products

Many firms simply buy such raw materials directly from companies in the foreign countries. In other cases, the domestic firm will set up its own operation in the country. This reduces the risk of supply interruption because the domestic company has greater control of the foreign operation. Of course, other problems can arise that threaten such supplies, including war or government appropriation of the firm's property. These risks can be reduced by diversification, that is, by developing raw materials sources in several countries.

Extending Market Power

As discussed in chapters 9 and 10, some firms have a degree of market power that gives them some control over the price of their products and may allow them to earn economic profits. This power may arise because the large size of the firm enables it to produce at lower cost than smaller rivals (i.e., economies of scale are present); the firm may have proprietary technology that results in low-cost production or even the monopoly production of a product; and/or the firm may have achieved a degree of product differentiation that allows it to sell above the price that would prevail in a competitive market. To some extent, the large multinational firms are able to use all three of these sources to increase their profits by expanding into foreign countries.

Consider the production of automobiles. Efficient production requires that large plants be used that incorporate assembly-line techniques to mass-produce cars. That is, there are significant economies of scale in this industry. Only very large firms are able to finance such facilities. Further, the production equipment and techniques are very sophisticated and require substantial technical expertise that is not generally found outside the staff of the firms in the industry. Finally, brand-name recognition also is very important; nameplates such as Chevrolet, Ford, Toyota, and Honda have come to have important meaning for most consumers and make it difficult for new entrants to gain consumer acceptance for their cars.

In recent years, Japanese auto producers such as Honda, Mazda, and Toyota have opened large manufacturing plants in the United States. All of these firms are large enough to build large plants that allow efficient production, they have the technical capability to organize and run those plants efficiently, and their products are differentiated. Thus, their location in the United States is a natural extension of their market power into the U.S. economy.

Comparative Advantage

Finally, firms locate facilities in foreign countries to capture the *comparative advantage* offered by that country. Comparative advantage refers to a situation where the opportunity cost of producing two goods differs between two countries. One country is said to have a comparative advantage in the production of those goods for which the opportunity cost is lower than in the other country. Consider the following hypothetical production data for two countries, Japan and the United States, which are assumed to be producing only cars and boats. If the United States used all of its resources to produce cars, assume that maximum production would be 100, and if it only produced boats, maximum production would be 200. In Japan, assume that maximum production could be 300 cars or 50 boats. That is, maximum production in each country is

	U.S.	Japan
Cars	100	300
	or	or
Boats	200	50

The opportunity cost of producing each product in each country in terms of foregone production of the other good is

	U.S.	Japan
Cars	2 boats	1/6 boat
Boats	1/2 car	6 cars

Note that Japan has a comparative advantage in producing cars (i.e., each car costs only 1/6 boat compared to two boats in the U.S.) and the United States has a comparative advantage in producing boats (i.e., each boat produced in the United States costs 1/2 car compared to 6 cars in Japan). When the opportunity cost of production differs between two countries, both can be better off by specializing in the production of the one product where they have a comparative advantage and then by trading with the other country.

For example, assume that both countries produce both goods and there is no trade between them. One possible production combination is shown here:

	U.S.	Japan	Total
Cars	50	180	230
Boats	100	20	120

Because there is no trade, the production in each country also represents the amount consumed in that country. Now assume that Japan produces only cars and the United States produces only boats; the maximum production is now

	U.S.	Japan	Total
Cars	0	300	300
Boats	200	0	200

Note that total production in the two countries has increased by 70 cars and 80 boats compared to the original levels.

Consumers in the United States and Japan want both cars and boats, so boats are traded for cars and vice versa. One possible final consumption pattern is

	U.S.	Japan	Total
Cars	60	240	300
Boats	120	80	200

This means that of the 300 cars produced in Japan, 60 were exported to the United States and 240 consumed in Japan. The United States produced 200 boats, of which 120 were exported to Japan.

Now consumers in both countries have more of both goods than they did before specialization and trade. Economic welfare has been enhanced with no increase in the amount of labor and capital employed; these factors of production have simply been reorganized so that each country captures its comparative advantage.

Usually, one need not know anything about the amount of resources devoted to the production of any good to determine which country has the comparative advantage in producing a particular good. If a country has a large share of world production and exports much of its output, it is probable that that country has a comparative advantage. Japan appears to have an advantage in electronics equipment, while the United States has a comparative advantage in the production of agricultural products.

The principle of comparative advantage will determine where multinational corporations locate their production facilities. Particular products will be produced in those countries where the opportunity cost of production is low.

Case Study

Matsushita: The Ultimate Multinational Firm

Matsushita Electric Industrial Company, Ltd. is the largest producer of consumer electronics and industrial communications and information products in the world. Annual sales for this firm total about $62 billion. The firm was organized in 1935 by the late Konosuke Matsushita, whose corporate philosophy was based on the principle of "... contributing to the peace, happiness, and prosperity of mankind through the abundant supply of quality consumer goods at reasonable prices." If sales volume and annual payrolls are any indication, consumers and workers are happier and more prosperous because of his efforts.

The firm began the post-World War II period making washing machines and refrigerators. Later it developed new products, including color television sets, stereo equipment, air conditioners, microwave ovens, and a variety of other products for both consumers and industry. Its single biggest success was the videocassette recorder (VCR), which it pioneered along with the Sony Corporation. This product has changed the entire entertainment industry in the past 20 years and has generated enormous sales and profits for Matsushita. Consumers quickly recognize the firm's brand names, such as Panasonic, Technics, National, and Quasar. Often products sold by other firms under other brand names came from Matsushita factories.

The company is truly multinational in scope. Its 343 factories are scattered throughout the world, as are its 271,000 employees. The 205,000 stockholders in the firm bought their shares on markets in North America, Europe, Japan, and other parts of Asia. Until 1988, the company's accounting, finance, and marketing operations were controlled from headquarters in Japan, but recognition of the company's worldwide orientation led to a decentralization of these central functions to a variety of offices in the United States, Europe, and Japan.

In 1991, Matsushita made a major diversification move by acquiring MCA, Inc., a U.S. firm that is a major producer of movies, television shows, and recorded music. Not only are these products complementary with most of the company's electronics products, this acquisition allowed Matsushita to benefit from the comparative advantage of the entertainment industry in the United States. ■

SUMMARY

Selecting the location for a business is important because the costs of moving output, workers, and customers across space are significant. To economize on these costs, people and firms tend to locate together. The urbanization process observed in the United States reflects the economic advantages of concentrating people and business together.

A firm serving a linear market should locate at the median location, where there are an equal number of customers on both sides of the firm. Such a location will minimize the total transport costs associated with serving those customers. If two firms serve a linear market where buyers are evenly dispersed along the market and demand does not vary with the seller's location, both will tend to locate at the midpoint. If demand is inversely related to distance from the seller, the two vendors would tend to locate closer to the socially efficient points.

Firms that use raw materials from one source and sell finished output at another site choose a cost-minimizing location that depends on the cost of shipping raw materials relative to the cost of shipping output. Because of additional loading and unloading costs, firms tend not to locate at points intermediate between raw material sources and markets. Products that are weight losing in production tend to be produced at the raw materials site, whereas firms that produce weight-gaining products usually locate at the market.

The market areas served by competing firms depend on their relative production and transportation costs. Firms with lower production and/or transportation costs have larger market areas than their competitors. The cost differentials can become so great that the higher-cost firm loses its entire market area to the lower-cost firm.

Firms do not locate facilities in an area until the threshold level of demand has been achieved. This threshold is defined as the tangency of the average cost function and the demand curve. At that point, price covers average cost, and a normal profit is possible. Some activities have low threshold levels and are found in even the smallest towns. Conversely, other activities tend to be found only in larger urban areas.

The industrial location decision involves a comparison of the characteristics of the proposed facility to the locational attributes of the areas being considered. Location factors can be classified as primary or secondary. Primary factors include geographic location, skilled labor availability, the cost and availability of energy resources, and availability of land transportation. Secondary factors influencing location include an area's physical environment and government attitudes and policy. Usually, differences in state and local taxes are of secondary importance in selecting an industrial location. Generally, firms select that location that will maximize the present value of future profits.

Firms locate plants and other facilities in foreign countries to provide dependable sources of raw materials, to extend market power, and/or to capture the comparative advantage of producing in those countries. A country has a comparative advantage if its opportunity cost of producing a product is lower than it is for other countries.

Discussion Questions

17-1. Why do some businesses such as automobile dealers tend to locate together?

17-2. What location factors would be of primary importance and which would be of little significance to a firm in each of the following industries: retailing, oil refining, higher education, and lumber?

17-3. Is it wise for a local government to subsidize industries that locate in its city? What advantages and disadvantages are associated with such action?

17-4. In the 1970s and 1980s, technical and vocational education programs were expanded rapidly in states that were aggressively seeking new industry. What connection, if any, do you see between such education programs and the industrial location decision?

17-5. What is the market area for your university or college? That is, from what geographic area do most of the students come? Did the location of the school influence your enrollment decision? Explain.

17-6. What would you estimate to be the market area of each of the following firms?
   a. General Motors Corporation
   b. A soft-drink bottling plant
   c. Chase Manhattan Bank
   d. First National Bank of Manhattan, Kansas

17-7. Consider your community and two others that you are familiar with—one half as large and one twice as large (if there is one). What goods and/or services are available in your city that are not available in the smaller one? What is not available in your city that is found in the larger city? What is the explanation for this distribution of economic activities?

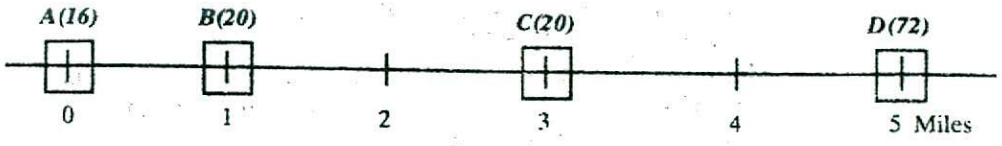
17-8. In the chapter, computers and automobiles are used as examples of products for which firms tend to produce a median product. What other goods and services have similar median product characteristics? Explain.

Problems

17-1. Customers of a firm are located along a road as indicated by the letters *A* through *J* in the following figure.

To service these customers, two truckloads of material must be delivered to each customer every week. The cost of making a delivery is $1.50 per mile.
   a. Determine the cost-minimizing location for the firm.
   b. Determine the total transportation cost each week assuming that (1) the firm locates at the median location; and (2) the firm locates at the mean location.

17-2. United Express makes weekly deliveries to each of a number of customers located in cities along an interstate highway. The location of these cities, *A*, *B*, etc., and the number of customers in each is shown in the following figure.



On one trip, four customers can be serviced. The transportation cost is $1 per mile.

a. Determine the median location for locating the company's delivery center.
b. Determine the total transportation costs at the median location.

17-3. Suppose that 500 swimmers are evenly distributed at intervals of 2 feet along a beach 1,000 feet long (i.e., one at the 2-foot mark, one at the 4-foot mark, etc.). Initially assume that each will buy one hot dog during the day from the nearest vendor.

a. If two hot dog vendors, *A* and *B*, locate at each of the following distances from the end of the beach, how many hot dogs will each sell during the day and what will be the total distance walked by the customers?

	Location of:	
Alternative	***A***	***B***
I	200 ft	400 ft
II	500 ft	500 ft
III	250 ft	750 ft

b. Repeat the exercise of part (a) under the assumption that demand is inversely related to distance. Specifically, assume the following demand schedule for hot dogs:

Distance of Swimmers from Nearest Vendor	***Number of Hot Dogs Purchased***
Within 100 ft	3
101 to 200 ft	2
201 to 300 ft	1
More than 300 ft	0

17-4. U.S. Cement, Inc., uses 2 tons of raw material (i.e., limestone, clay, coal, and gypsum) to manufacture one ton of cement. All of the raw materials are available at site *A* and all final output is sold in a market (*B*) located 20 miles away. The transportation cost per ton of raw materials is $0.40 per mile, while the per-ton cost of transporting cement is $0.60 per mile. Assume that rail is the only transportation alternative.

a. Determine equations for the raw materials, output, and total transportation cost functions and plot these on a graph.
b. Determine the location that minimizes total transportation costs (i.e., site *A*, *B*, or an intermediate location).

17-5. Consider two competing plants located at points *A* and *B* that are 10 miles apart. Determine and sketch an equation for the market boundary separating these plants for each of the following cases:

a. Product price is equal at both plants (i.e., $P_A = P_B = 20$) and per-mile transport costs per unit of output from each plant are the same (i.e., $s_A = s_B = 2$).
b. $s_A = s_B = 2$ but $P_A = 25$ and $P_B = 20$.
c. $s_A = 2, s_B = 3, P_A = 20$, and $P_B = 24$.

17-6. Mario's Fine Foods, a chain of ethnic restaurants, is considering locating a restaurant in Millville and one in Yorktown. The total cost and marginal cost functions of one of their restaurants are

$$TC = 1{,}000 + 2Q + 0.005Q^2$$

and

$$MC = 2 + 0.01Q$$

where TC is total cost, MC is marginal cost, and Q is the number of meals served per week. The following demand and marginal revenue functions for each city have been estimated using data from market surveys:

MILLVILLE: $P = 10 - 0.005Q$, $MR = 10 - 0.01Q$.

YORKTOWN: $P = 6 - 0.005Q$, $MR = 6 - 0.01Q$.

a. Determine an equation for the average cost function. Sketch the average cost function and the two demand functions on the same graph. Use selected points from $Q = 0$ to $Q = 1{,}000$.

b. What output rate in terms of meals per month would minimize the average cost per meal?

c. Using the principles of threshold analysis, is a Mario's restaurant feasible in Millville? Yorktown? Explain.

d. For either or both cities where the restaurant is feasible, determine the profit-maximizing price–quantity combination, total revenue, total cost, and total profit.

17-7. The maximum production rate of wheat and nuts (both measured in bushels) for each of two countries is shown here:

	A	*B*
Wheat	1000 or	200 or
Nuts	4000	1000

a. What is the opportunity cost of production of each good in each country? Which country has a comparative advantage in wheat and which in nuts? Explain.

b. If the countries specialize in producing the product for which they have a comparative advantage, determine the amount produced in each and the total production for both countries combined.

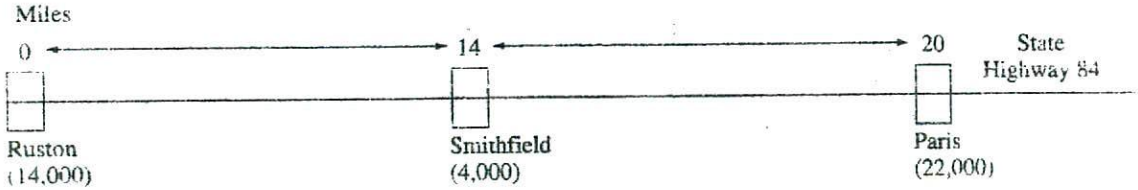
c. With specialization and trade, demonstrate that consumers in both countries can be better off than under a "no specialization—no trade" arrangement.

d. What are the implications of your analysis in parts (a)–(c) for the location of industry between these two countries?

Integrating Case Study VII

GMG Entertainment, Inc.

GMG Entertainment, Inc. operates movie theaters in small cities throughout the United States. In many markets, the firm's theater is the only one in the area and typically serves several communities. GMG is considering the location of a theater in a regional market in Texas that consists of three communities along State Highway 84. The location of these cities and their population are shown below. In each city, 25 percent of the population are children under the age of 14.



The number of movies demanded each month for each 100 persons in a city is given by the relationships

$$\textit{ADULTS:} \quad Q_A = 100 - 10P_A - 1.5\,D$$
$$\textit{CHILDREN:} \quad Q_C = 100 - 15P_C - 2\,D$$

where Q_A = the number of adult admissions per month per 100 population; Q_C = the number of children's admissions per month per 100 population; P_A and P_C are the prices of an adult and a child's admission; and D is distance from the customer to the theater in miles. Thus, if a_i represents the population (in hundreds) in the ith city, the monthly demand functions for that city would be:

$$Q^i_A = a_i\,(100 - 10P_A - 1.5D)$$

and

$$Q^i_C = a_1(100 - 15P_C - 2D)$$

For example, in Ruston, where there are 140 (hundred) adults, the demand function for adults would be

$$Q^R_A = 140(100 - 10P_A - 1.5)$$

The cost of building the theater is estimated to be $450,000, which includes the following:

Land: $50,000
Construction: $300,000
Equipment: $100,000

The board of directors of GMG insists that all capital investment projects be evaluated using 16 percent as the opportunity cost of capital. It is anticipated that the theater would have a 15-year life. GMG's marginal tax rate on profit is 40 percent.

In addition to having overall responsibility for the theater, the manager also operates the projection equipment. She works seven nights per week. The jobs of ticket seller/taker and refreshment sales are filled by three workers whose schedules are varied so that two are on duty each night. It has been determined that janitorial service and utility expense (primarily for heating) vary with the number of customers and are estimated to total $0.15 per admission.

Estimated costs for operating the proposed theater include the following:

Fixed Costs		
Movie rentals	$51,600	per month (plus a variable cost of $0.10 per admission)
Property taxes	$500	per month
Manager	$1,800	per month
Employees	$2,100	per month (total for all three employees)

Variable Costs		
Utilities, janitorial services, etc.	$0.15	per admission

Thus, marginal cost is $0.25 per admission, and fixed costs each month are $56,000.

Requirement

Based on the information provided, prepare a report that determines the profit-maximizing location for the theater and the optimal prices for children and adults. Because the theater must have access to such urban services as police and fire protection, it must be located in one of the cities. Include in this report an evaluation of whether this capital investment should be made. Assume that there will be no change in demand or costs over the period, and, for simplicity, consider all revenues and operating costs as being collected or incurred at the end of each year.

PART VIII

Business Decisions and Government

CHAPTER

18 Taxes and Decision Making

PREVIEW

A large business may be required to pay dozens of different taxes and fees. At the federal level, there is the corporation income tax, the employer's contribution to the social security or payroll tax, and various excise taxes. The state government is likely to utilize a sales tax, excise taxes, its own corporation income tax, and a payroll tax used to pay unemployment compensation. At the local level, firms may face property taxes, license fees, and other assessments.

In addition to providing revenue to finance the activities of government, the various taxes have something else in common—each can affect managerial decision making. Taxes affect managerial decisions in several ways. Sometimes tax considerations determine the legal form selected for a business enterprise. For example, a desire to avoid payment of the federal corporation income tax may cause a firm to continue to operate as a partnership rather than to incorporate. Taxes can affect methods of doing business. Effluent taxes (i.e., taxes on pollution) provide incentives to change production methods, and property tax rates may affect location decisions. Taxes can also be an important determinant of the demand for a product sold by a business. For example, for many years an excise tax imposed on margarine made it difficult for that good to compete with butter. In contrast, a temporary income tax credit available to the purchasers of solar heating systems allowed such equipment to become more competitive with other methods of heating.

This chapter uses the tools of economic analysis developed in previous chapters to analyze the ways that taxes affect managerial decisions. First, the impact of an excise tax on equilibrium price and quantity is discussed. Next is a consideration of optimal strategies when a profits tax is imposed. The third section examines how taxes on inputs affect production decisions, and the fourth analyzes the effect of the property tax. The chapter concludes with an evaluation of preferential tax treatment, such as the income tax deduction for interest expenses, investment tax credits, and accelerated depreciation.

EXCISE TAXES

An *excise tax* is a sales tax levied on a particular good or service. For example, the federal government imposes excise taxes on gasoline, cigarettes, and liquor. Most states tax these three commodities, and many utilize excise taxes on other goods and services, such as soft drinks, hotel lodging, and theater tickets. In addition to providing revenues, excise taxes are sometimes used to decrease the quantity demanded for a product by increasing its effective price. Substantial excise taxes levied on cigarettes and liquor have always been justified as a means of reducing consumption of those products. Similarly, taxes on gasoline are used as a means of encouraging energy conservation. This approach is common in Europe, where excise taxes on motor fuel may cause retail prices to be double or triple the wholesale price of gasoline.

Case Study

Taxing Beer in Thailand and Gasoline in Washington, D.C.

Using excise taxes to alter consumer behavior can be a tricky undertaking. Sometimes the results are not as intended. Recently, the government of Thailand imposed a heavy excise tax on beer. The tax was designed to raise revenue and to reduce beer drinking. As a consequence of the tax, the price increased from $1.25 to $1.75 per pint and beer consumption dropped by 50 percent. But because the tax applied only to beer, other alcoholic beverages became relatively less expensive. For example, a pint of whiskey could be purchased for about the same amount of money as a pint of beer. The result was increased consumption of substitute beverages such as whiskey and other hard liquors. Thus, by taxing beer, the government inadvertently encouraged people to switch to beverages with a higher alcohol content. Consequently, the consumption of alcohol may actually have increased as a result of the tax.

In 1980, the city of Washington, D.C., imposed a 6 percent excise tax on all gasoline sold within the District. Because the elasticity of demand for gasoline is low, it was expected that the tax would generate substantial revenues, which would help the city reduce a large budget deficit. But 6 months later, the tax was repealed. What happened?

Washington, D.C., is only 10 miles square. Rather than pay the gasoline tax, motorists on the edges of the District simply bought their fuel at lower prices in the suburban areas of Maryland and Virginia, which surround the city. During the first month the tax was in place, the amount of gasoline sold in Washington, D.C., dropped by about one-third. Consequently, the anticipated tax revenues did not materialize, and D.C. gas station owners lost sales and profits. ■

Typically, it is the ultimate responsibility of the seller to pay an excise tax. Consider a federal excise tax of 20 cents per pack on cigarettes. A store that sells cigarettes must remit 20 cents to the federal government for each pack sold. Now suppose that the tax is increased to 50 cents per pack. The store may raise the price of cigarettes by the amount of the tax increase, maintain the price at the pretax level, or increase the price by a part of the tax increase. The pricing policy used by the store does not matter to the government as long as the seller pays 50 cents in tax for each pack of cigarettes sold.

How should the store determine the price to be charged for cigarettes? A first reaction might be that the choice is obvious; the price should be increased by the full amount of the tax. But that would not necessarily be a wise decision. In fact, it might be a very poor choice under certain conditions. Basically, proper pricing policy requires consideration of the demand-and-supply functions for the product being taxed.

Price	Quantity Demanded	Quantity Supplied (before tax increase)	Quantity Supplied (after tax increase)
$2.55	22	10	6
2.70	20	12	8
2.85	18	14	10
3.00	16	16	12
3.15	14	18	14
3.30	12	20	16
3.45	10	22	18

Suppose that before the excise tax increase, the retail demand and supply schedules for cigarettes are as shown in Table 18.1. Note that supply equals demand at $3.00 per pack. Thus, the equilibrium quantity is 16 billion packs and the equilibrium price is $3.00.

Now consider the 30-cent increase in the excise tax from 20 cents to 50 cents. Let the prices shown in Table 18.1 be the total amount received by the seller. At each price level, the seller must now pay an extra 30 cents in tax. Hence, the net amount retained by the seller at any price level is 30 cents less than before. For example, if the consumer pays $3.00, from the perspective of the seller the price is the same as a price of $2.70 before the tax increase. But sellers' decisions are based on the amount of money actually received. Thus, at a price to consumer of $3.00, only 12 billion packs per year will be supplied. Similarly, a price of $3.15 generates the same supply response as did a price of $2.85 before the tax increase. Thus, the quantity supplied will be 14 billion. The other entries in the fourth column of Table 18.1 are determined in a similar manner.

The second and fourth columns of Table 18.1 can be used to determine the new equilibrium price. Note that supply and demand are equal at $3.15 per pack. This equilibrium price is 15 cents higher than the equilibrium before the 30-cent tax increase. The implication is that the tax increase is being shared by consumers and sellers. Consumers are paying 15 cents more for each pack of cigarettes and sellers are paying the other 15 cents. As expected, the new equilibrium quantity (14 billion packs per year) is less than the previous equilibrium amount (16 billion).

Figure 18.1 is a graphical representation of Table 18.1. The supply response to the tax increase is shown by shifting the supply curve up by the amount of the tax (30 cents). This shift (shown in the figure at $S'S'$) reflects the increased cost to the firm of selling cigarettes. The new supply curve indicates that less is supplied at any price than before the tax increase. The new equilibrium price of $3.15 occurs at the intersection of DD and $S'S'$.

In general, an excise tax is not borne equally by consumers and sellers. The actual impact of the tax depends on the relative slopes of the supply-and-demand curves of the product being taxed. Consider Figure 18.2a. In this case, the demand curve is vertical, indicating that demand is totally inelastic. Before the increase in the excise tax, the equilibrium price is $3.00. After the increase, the new equilibrium price is $3.30, while the quantity remains unchanged. Note that the price has risen by the full amount of the tax increase. There is no sharing of the tax increase; consumers pay the entire 30 cents.

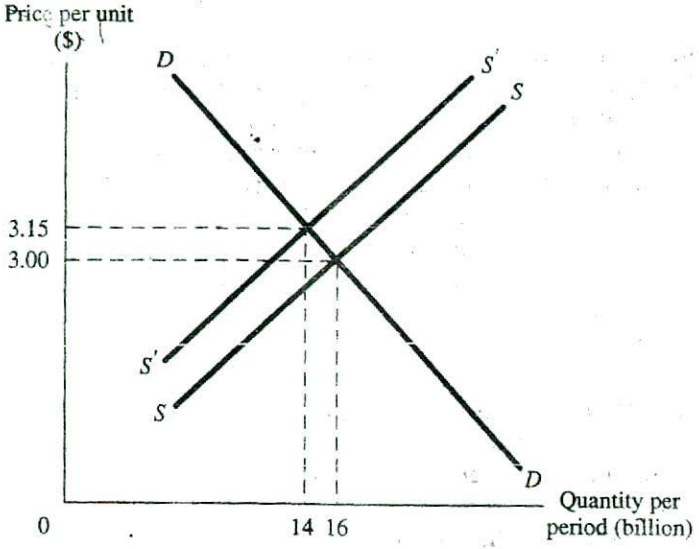
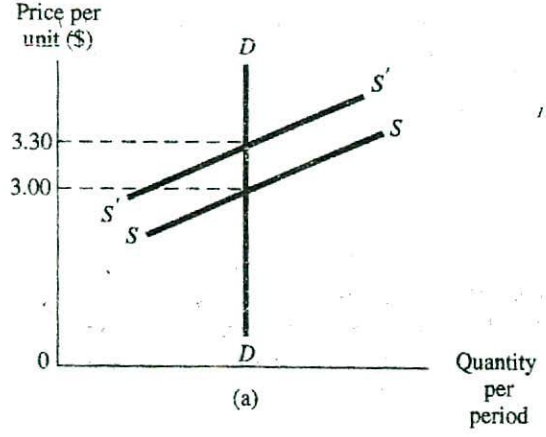
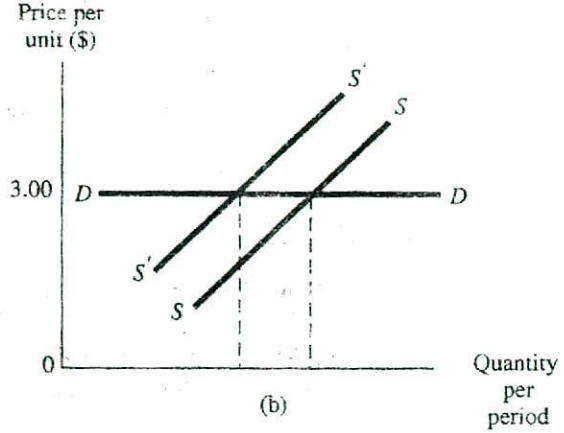
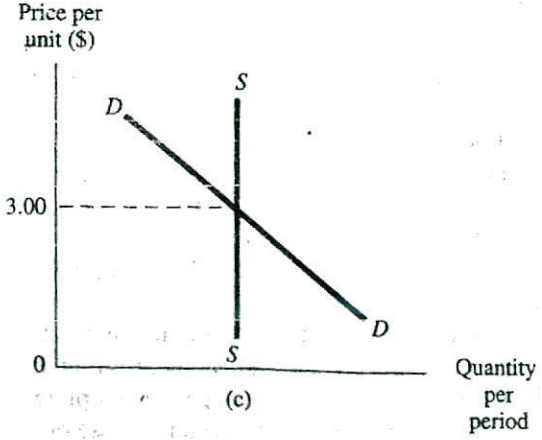


FIGURE 10.1 An Excise Tax Shifts Costs to Consumers and Sellers







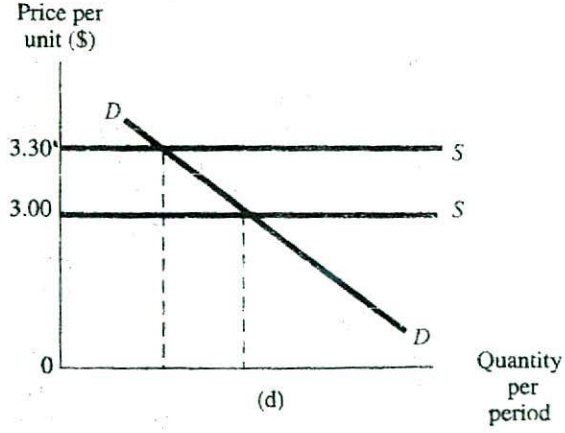


FIGURE 10.2 Incidence of the Excise Tax Limiting Cases

The explanation is that quantity demanded is totally insensitive to price. The quantity demanded at a price of $3.00 per pack is the same as $3.30 per pack. Thus, the entire tax increase can be passed on to the consumer with no loss of sales to the producer.

Figure 18.2b depicts an upward-sloping supply curve and a horizontal (perfectly elastic) demand curve. As a result of the tax increase, the quantity decreases, but the equilibrium price does not change. Thus, the seller pays all of the tax. The reason is that because demand is totally elastic, any price increase will cause quantity demanded to be zero. Hence, no portion of the tax increase can be passed on to the consumer. The entire tax must be absorbed by the seller.

Figures 18.2c and d show the incidence of an excise tax for vertical and horizontal supply curves. If supply is totally unresponsive to price (Figure 18.2c), there is no shifting of the supply curve in response to a tax increase. Equilibrium quantity remains unchanged, and the entire tax is paid by the seller. In contrast, if the supply curve is horizontal (Figure 18.2d), consumers will bear the entire burden of the tax. This is because any decrease in the revenue per unit retained by the seller would cause quantity supplied to be zero. Thus, the price increases by the full amount of the tax change.

In general, the effect of an excise tax on the price of a good or service depends on the elasticities of supply and demand. As demand becomes more elastic, a greater proportion of the tax must be paid by sellers. Similarly, as supply becomes more elastic, a larger share of the tax will be passed on to consumers. Thus, if managers have information on elasticities of supply and demand, they will be able to anticipate the effects of changes in excise tax policies and plan accordingly. Although precise estimates of elasticities are not always available, even a rough guess can aid managers in making pricing decisions in response to changes in taxes.

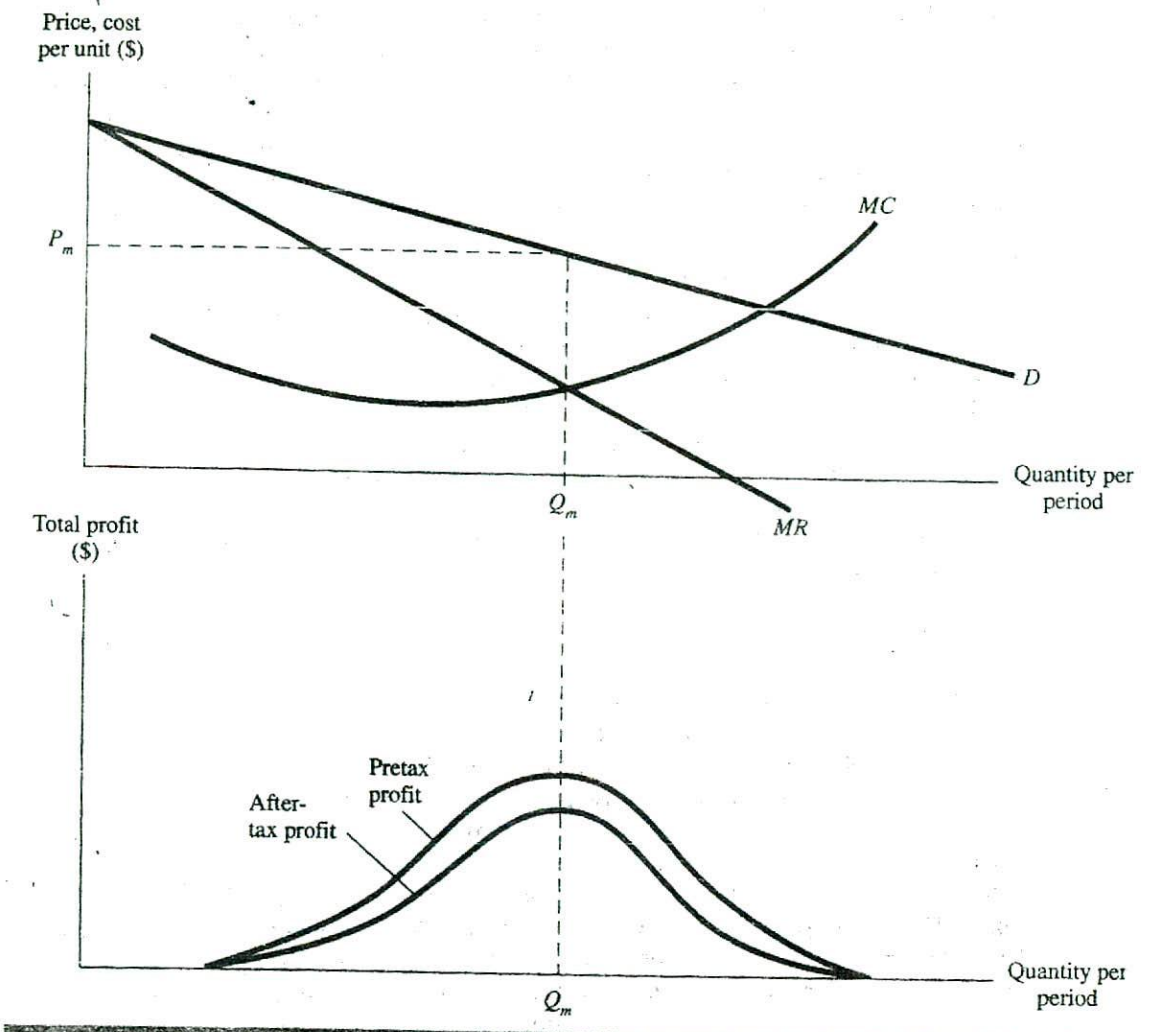
Key Concepts

- As demand becomes less elastic, the consumer pays a larger proportion of an excise tax. With completely inelastic demand, the entire tax is paid by the consumer.
- As supply becomes less elastic, sellers pay a larger proportion of an excise tax. With completely inelastic supply, the entire tax is paid by the seller.

TAXES ON PROFIT

In chapter 9 it was shown that in the long run, firms in perfectly competitive markets earn no more than a normal rate of profit. In contrast, firms with market power may earn substantial economic profit. This economic profit represents a transfer of wealth from consumers to the shareholders of the firm. One means of redistributing this wealth is to impose a tax on profits. Ideally, such a tax would be paid only by firms that earn economic profits. As a practical matter, it is not possible to identify precisely the amount of economic profits being earned. Thus, such taxes usually are structured as a percent of total accounting profit earned by a firm.

The purpose of this section is to analyze how a tax on profit affects pricing and output decisions of managers. It will be shown that the impact of such a tax depends on the objectives of managers. Specifically, it will be demonstrated that managers attempting to maximize total revenue respond to a profit tax differently than managers whose objective is profit maximization.



Profit Maximization and Profit Taxes

The profit-maximizing price and quantity for a monopoly are shown in Figure 18.3. Also shown is total profit earned at various output rates. Note that profit increases to output rate Q_m and declines for higher output rates.

Now assume that a proportional tax of t percent is levied on each dollar of profit. After-tax profit is also shown in Figure 18.3. Note that the effect of this tax is to reduce total profits by t percent at each output rate. However, it should also be observed that although profit is less than before the tax, the output rate that results in maximum profit has not changed. The firm maximizes after-tax profit by producing Q_m units as before. Because the profit-maximizing quantity is the same, the price set by the firm for the product will not change either. The profit-maximizing price is still P_m. In the short run,

the implication is that the entire tax is absorbed by the firm in the form of reduced profits. None of the tax is passed on to consumers in the form of higher prices.

However, a tax on profit can affect prices and output in the long run. Such a tax reduces the net return on investment. This means that less investment will take place, and hence the amount of capital in an industry will be reduced. A result is that productive capacity will increase less rapidly than if the tax had not been imposed. Thus, because quantity supplied will be less, in the long run prices are likely to be higher because of the tax on profit.

Revenue Maximization and Profit Taxes

The models of firm behavior developed thus far have assumed that profit maximization is the goal of managers. However, profit maximization is not the only approach to developing models of the firm. Alternative assumptions regarding objectives of managers can be used. One alternative is revenue maximization, which presumes that a manager acts to maximize total revenues received by the firm. Use of the revenue-maximization assumption can result in conclusions that are quite different from those obtained by assuming profit maximization.

Figure 18.4 shows that revenues are maximized at Q_1, where marginal revenue is zero. This result follows from the definition of marginal revenue—the extra revenue that results from producing one more unit of output. For any output rate less than Q_1, additional output increases total revenue because marginal revenue is positive. But beyond Q_1, marginal revenue is negative and hence total revenue declines.

The lower panel of Figure 18.4 shows profit at each rate of output. Let π_0 represent a target return set by management. Suppose that this target represents the minimum profit acceptable to the owners of the firm. Note, however, that profit is less than π_0 at Q_1. Thus, managers must act to increase profit to at least the target return. In recognition of this constraint, a modified version of the revenue-maximizing hypothesis may be considered. This variant assumes that the objective of managers is to maximize revenue subject to the constraint that the firm earn at least its target rate of profit. Output rate Q_2 in Figure 18.4 is the quantity that meets this objective. Any output greater than Q_2 yields less than the target profit. Because marginal revenue is positive, output rates less than Q_2 generate reduced total revenue. Corresponding to the output rate Q_2 is the price P_2.

Now suppose that a proportional profit tax is imposed. Figure 18.4 shows that after-tax profit at Q_2 is now less than π_0. Thus, to maximize revenue while satisfying the target profit constraint, output must be reduced to Q_3 and the price increased to P_3.

Note how the effect of the profit tax depends on the assumed objective of managers. For profit maximization, there is no short-run change in price or quantity as a result of the tax. In contrast, a revenue-maximizing manager subject to a target profit constraint will reduce output and increase price in the short run in response to a higher tax on profit.

Key Concepts

- If the firm's objective is profit maximization, the optimal price and quantity in the short run will remain unchanged if a profit tax is imposed. However, in the long run, a profit tax may cause firms to reduce their investment in productive capacity.
- For revenue-maximizing managers subject to a target profit constraint, a profit tax will result in reduced output and higher prices.

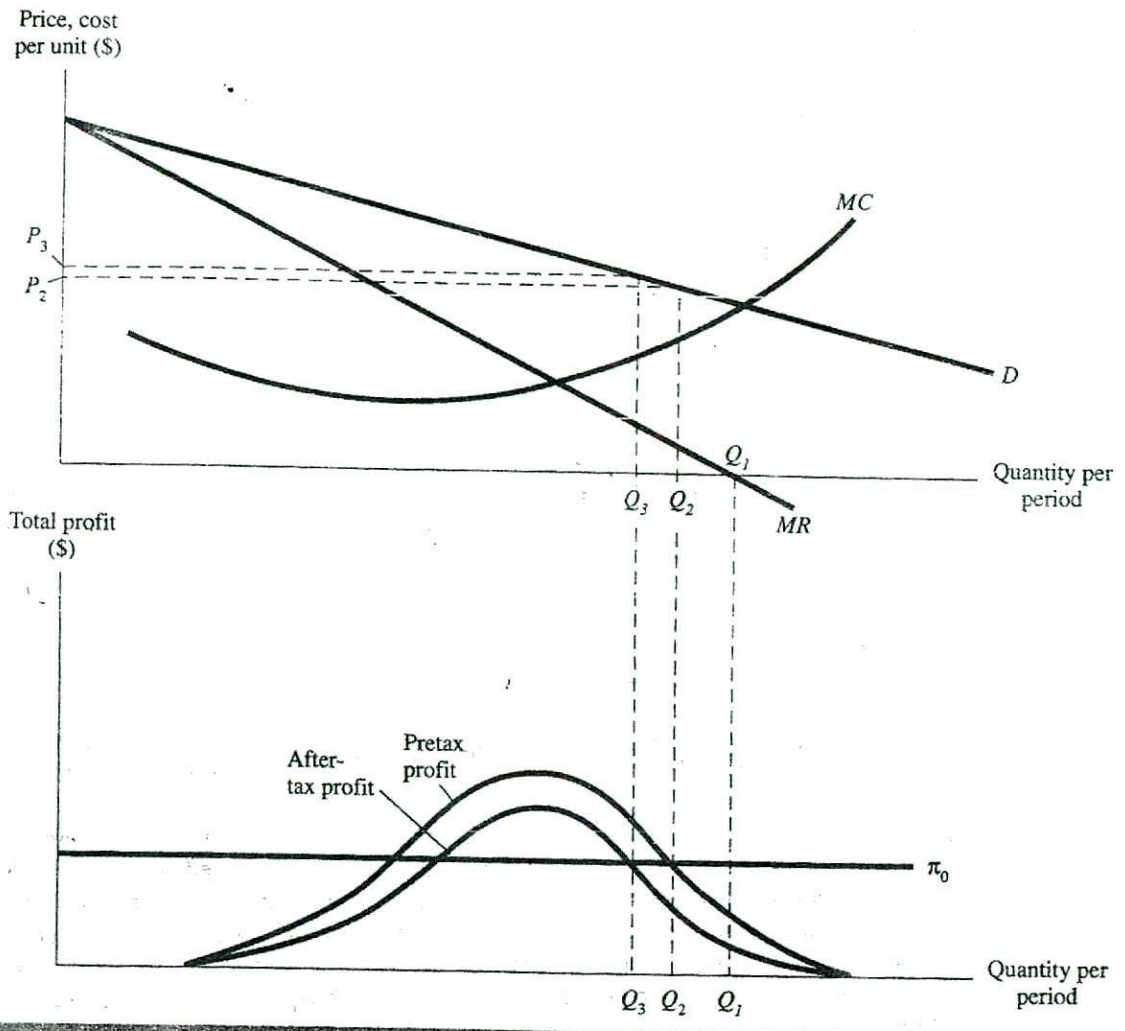


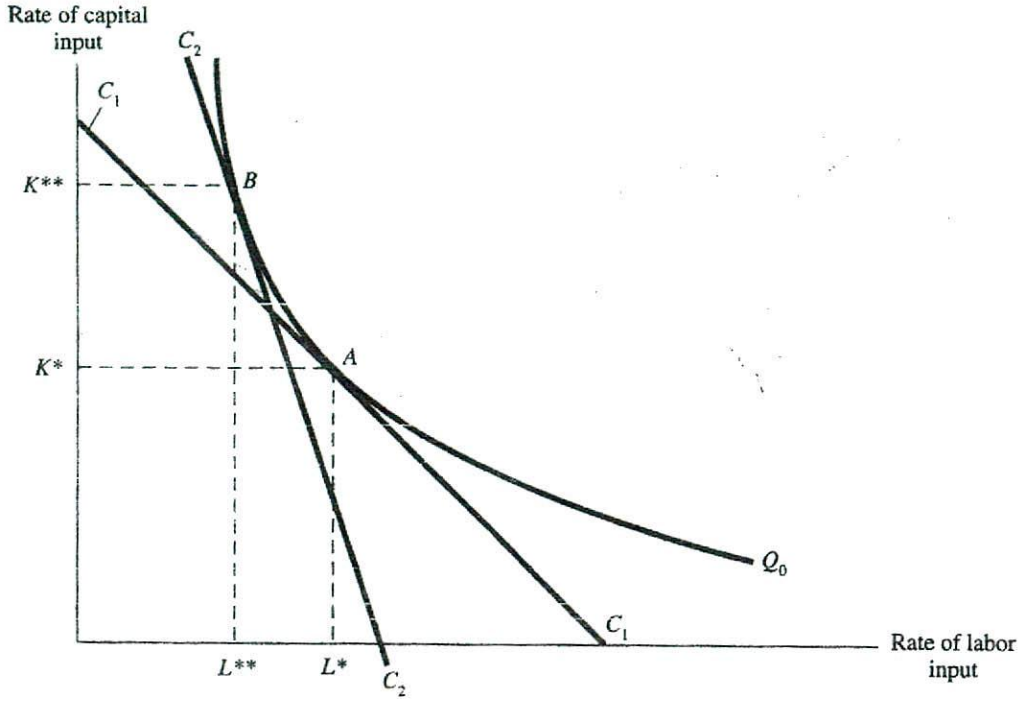
FIGURE 18-2 Response to an Input Tax

TAXES ON INPUTS

A firm may be taxed or required to make other payments to government based on the amounts of certain inputs utilized in its production process. One example would be a state excise tax on energy use. Another is money paid (based on the firm's total wage bill) to unemployment compensation or disability funds for those workers who lose their jobs or are injured while at work. An effluent tax based on emission levels of certain pollutants is a third example of a tax on inputs. It can be thought of as a tax on the use of water and air as depositories for waste materials.

Cost Minimization and Input Taxes

When a tax is levied on an input, information is provided to managers that relative prices of inputs have changed. Specifically, managers are signaled that the price of the taxed in-



put is relatively higher than before the imposition of the tax. If managers are attempting to minimize costs, the tax may cause a change in the mix of inputs used for production.

Consider a firm that uses two inputs, capital and labor, to produce its product. Assume that the objective of managers is to minimize the cost of producing any given rate of output. Figure 18.5 depicts an isoquant for Q_0 units of output. As discussed in chapter 6, this isoquant shows all the efficient combinations of capital and labor that can be used to produce Q_0.

Let w be the cost of labor and r be the cost of capital. The isocost line C_1 shown in Figure 18.5 indicates the combinations of capital and labor that can be purchased for a given total expenditure, C_1. The slope of this isocost line is $-w/r$. For input prices w and r, the most efficient combination of capital and labor to produce Q_0 units is at point A, where the isocost line C_1 is tangent to the isoquant. Point A specifies use of L^* units of labor and K^* units of capital.

Now suppose that a tax of $\$t$ is imposed on each unit of labor used by the firm. What will be the least-cost combination to produce the specified rate of output, Q_0? The tax raises the price of labor to $(w + t)$ and changes the slope of the isocost line to $-(w + t)/r$. As before, the cost-minimizing rates of capital and labor are determined by the point where the lowest isocost line is tangent to the isoquant. The lowest isocost line is now C_2. Thus, the optimal combination of inputs to produce Q_0 is at B, where C_2 is tangent to the isoquant. Point B indicates K^{**} units of capital and L^{**} units of labor. This combination includes more capital and less labor than before the tax. Thus, the impact of a tax on labor is that managers substitute capital for labor in the production process.

This result is true in general. If technology allows substitution between inputs, a tax imposed on one input will cause greater use of the other inputs in the production process. One implication of this result is that input taxes can be used as a tool for public policy. For example, a tax levied on energy use would encourage energy conservation.

Effluent Taxes

If private costs of production are less than the total costs imposed on society, too many resources may be allocated to an activity. The production of paper is a good example. A paper mill generates residuals in the form of chemicals, particulates, and odors that are discharged into the environment. The high level of pollution resulting from paper production occurs because the firm does not consider all the costs of its activities. The problem is that the prices of some of the inputs used by the firm do not reflect their true social cost. Specifically, the air and water used for waste discharge may be considered free goods by the managers of the firm.

Excessive levels of pollution occur because of the lack of markets for air and water used in the production process. A business that uses energy or labor must pay for these inputs. But no one can exert an effective claim to ownership of the air, oceans, large lakes, and important rivers. Because there are no well-defined property rights, some individuals and firms can make free use of these resources. Further, there is no market mechanism to force them to take into account the social costs of their actions.

One approach to the problem of excess pollution is the use of effluent taxes. Under this approach, firms are allowed to choose the level of emissions but are taxed for the privilege. A tax on pollution can be considered as a tax on the use of water and air. By raising the price of water and air in this use, policymakers encourage managers to substitute other inputs in the production process. If a tax on emissions is set high enough, a firm will find it can reduce costs by installing pollution-control equipment that reduces the amount of emissions.

An advantage of effluent taxes is that they encourage the efficient allocation of resources. First, the divergence between private and social costs can be reduced or eliminated because the firm is forced to take into account the total social cost of its actions. Second, by allowing firms to select the optimal trade-off between polluting and paying taxes, variations in the cost of emission abatement can be taken into account. Firms with high control costs may find it profitable to make smaller reductions and pay larger amounts of effluent taxes. These funds can be used to compensate those who are affected by the firm's pollution. In contrast, firms that can easily reduce their polluting may find it less costly to make substantial reductions in discharges and thus avoid the tax.

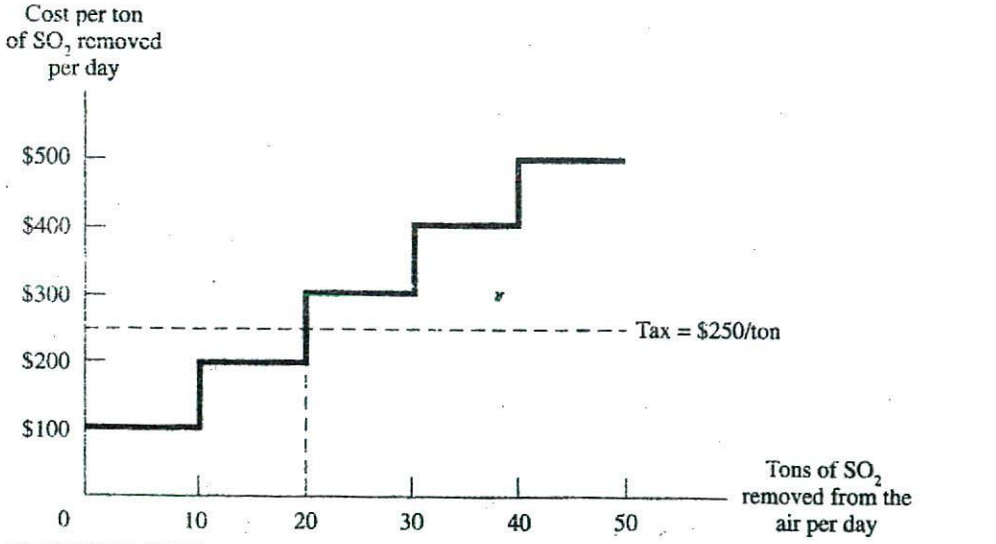
Although they have promise, effluent taxes are not a complete solution. It can be very difficult to determine the optimal rate of the tax. Theoretically, the marginal tax rate should be equal to the marginal benefit of pollution abatement. Thus firms would devote resources to reducing emissions until the marginal cost exceeded the marginal benefit. In actual practice, marginal costs are very difficult to determine, and any assessment of marginal benefits must be rather speculative. Effluent taxes have been used in Europe but have not found wide acceptance in the United States.

Example An Effluent Tax on Sulfur Dioxide Emissions

A community is experiencing serious air pollution resulting from emissions of sulfur dioxide (SO_2) from heavy industry in the area. The cost of reducing the emissions is not

the same for all polluters. For some firms the cost is relatively low, while for others it is quite high. Assume that there are five sources of SO_2 pollution—*A, B, C, D,* and *E*—and that each emits 10 tons per day. Costs per ton of SO_2 removed from the air are assumed to be constant (i.e., for each polluter the first ton removed costs the same as the last ton) and are \$100, \$200, \$300, \$400, and \$500, for sources *A, B, C, D,* and *E,* respectively. How could an effluent tax be used to reduce sulfur dioxide air pollution by 20 tons per day?

Solution The following figure illustrates the effect of a tax on SO_2 emissions. If an effluent tax were imposed, each pollution source would have the option of spending the money to eliminate pollution or paying the tax. Thus, the decision faced by a profit-maximizing manager would be to select the least-cost option. For example, if the tax were set at \$250 per ton, it would be cheaper for sources *A* and *B* to make the changes to eliminate their pollution. In contrast, for *C, D,* and *E* it would be less expensive to continue to pollute and to pay the tax. The net result would be a reduction of SO_2 emissions of 20 tons per day and \$7,500 per day collected from effluent taxes (i.e., 30 tons of pollution per day from *C, D,* and *E* times \$250 per ton). The tax revenue would be available to reduce pollution further or to compensate those affected by the pollution.



The Effect of a Tax on Sulfur Dioxide Pollution

OPERTY TAXES

Property taxes generate most of the revenues for local governments. Typically, the property tax is levied on land, structures, machines, and vehicles owned by a business. But there is considerable variability in property tax rates from one locality to the next. In one section of the United States the annual property tax assessment may represent about 1 percent of the market value of the property. In another location the property tax may be as high as 3 percent of market value. These variations in property tax rates can affect the value of the firm and also the decisions of managers. Such effects are the focus of this section.

Fixed Property

Certain types of property are essentially fixed in location. Land and permanent structures, such as large buildings, are the best examples. Once constructed, it would be extremely costly and difficult to move a large manufacturing facility or an office building to a different location.

Assume that an economy has two taxing jurisdictions, *A* and *B*. Consider a parcel of land in *A*, which has an exact counterpart in *B*. With no property taxes, suppose that both pieces of land have a market value of $1,000,000 and generate annual rents of $100,000. Thus, the rate of return of each parcel is 10 percent.

A $50,000 tax is levied on the property in jurisdiction *A*, but no property tax is assessed in jurisdiction *B*. As a result, the owner of the land in *A* must pay $50,000 in taxes every year. This means that net income for that property will be $50,000 and the after-tax rate of return will decline to 5 percent.

Now suppose that the two parcels of land are listed for sale at their pretax market values of $1,000,000. Clearly, the land in *B* would be preferred because of its higher rate of return. For the land in *A* to be competitive, its price would have to decline substantially. In fact, if the taxing policies of the two jurisdictions are not expected to change, the price would have to drop to $500,000. At this price the $50,000 in after-tax income of the property in *A* would provide the same 10 percent rate of return as the untaxed property in jurisdiction *B*.

When taxes are reflected in the value of property, the tax is said to be *capitalized*. The degree of capitalization of a tax is a function of the efficiency of capital markets in equalizing the rates of return on different assets. It is also dependent on the extent that the tax is perceived as permanent. For example, if the 5 percent tax is expected to be eliminated after 1 year, the effect on property values would be much less than if the tax is expected to be maintained at the 5 percent rate indefinitely.

Basically, the effect of taxes on land and other property that cannot be moved is to cause a one-time reduction in the value of the property. Hence, the full impact of the tax will fall on the owner of the property at the time the tax is imposed. Because any subsequent owner will acquire the property at a reduced price, only the original owner will experience a loss of wealth.

Case Study

Proposition 13 and Property Tax Capitalization

In 1978, voters in California approved a statewide tax limitation initiative known as Proposition 13. The measure was intended to provide home owners with some relief from the high property tax rates that existed throughout the state. The key provisions of the initiative were (a) a 1 percent ceiling on the property tax rate that any local government could impose, (b) a rollback of assessed property values to their 1975 levels, and (c) the requirement that property taxes in an area could not be increased without approval by a two-thirds majority in a local election. The net effect of the initiative was to substantially reduce property taxes for many California home owners. Fortunately,

at the time, the state government had a large budget surplus, which was used to compensate local governments for the loss of property tax revenues. Consequently, there was no immediate reduction in the quality of services provided by cities, counties, and other local governmental units.

The concept of property tax capitalization suggests that the reduction in property taxes mandated by Proposition 13 should have increased the value of houses in California. In fact, if the property tax was fully capitalized, the price of a house should have increased by the present value of the tax cut. Is the evidence consistent with the theory?

A study by Rosen examined the effects of Proposition 13 in the San Francisco area.* He collected data on house prices and property tax payments before and after the passage of the initiative. Using regression analysis, Rosen estimated that the degree of capitalization was approximately 22 percent. That is, housing prices in San Francisco increased by about one-fifth of the property tax saving resulting from Proposition 13.

Economists have conducted dozens of studies of property tax capitalization. Although the exact percentages differ, almost all of these studies conclude that property taxes are partially, but not completely, capitalized. Typically, the finding is that the degree of capitalization is from 20 to 40 percent.**■

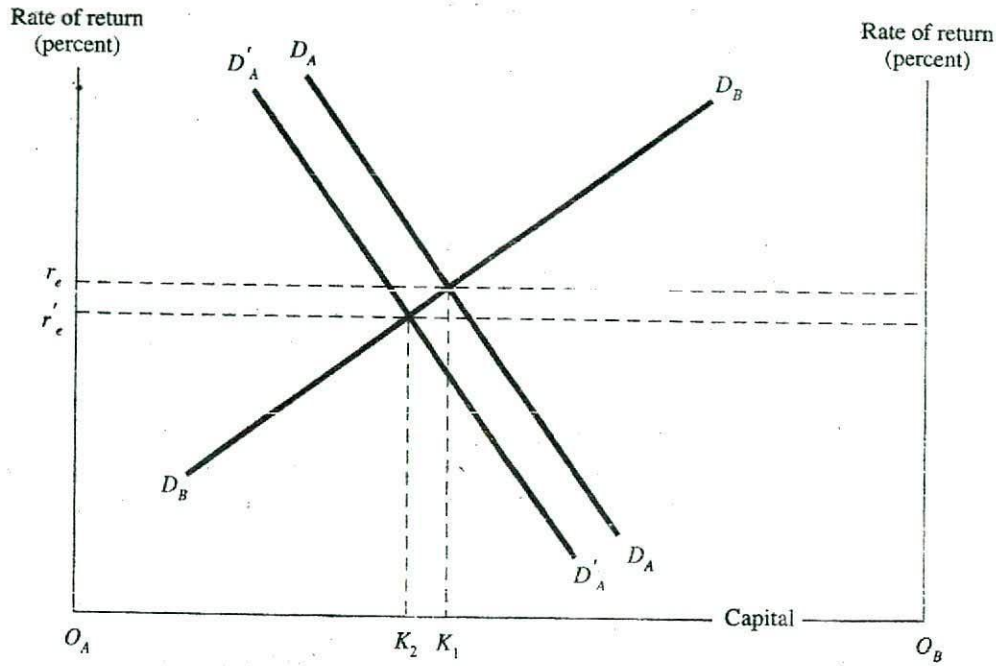
*K. Rosen, "The Impact of Proposition 13 on House Prices in Northern California," *Journal of Political Economy* 90 (1982): 191–200.

**For an excellent summary, see J. Yinger et al., *Property Taxes and House Values* (Boston: Academic Press, 1987).

Mobile Property

Some property can be moved from one location to another. For example, a large firm can easily transfer its vehicles among different offices. It may also be possible to relocate machines and other capital equipment. If there is a significant variation in property tax rates between taxing jurisdictions, the firm may have an incentive to move capital from the high-tax to the low-tax area.

The effect of this reallocation of capital is shown in Figure 18.6. Suppose that the firm has a fixed amount of capital, as measured by the length of the horizontal line from O_A to O_B. The relocation of capital from jurisdiction B to jurisdiction A is shown as a movement to the right along the horizontal axis. Similarly, the relocation of capital from A to B is shown by a leftward movement. For example, at O_A all capital is in jurisdiction B and at O_B all the firm's capital is in jurisdiction A. The curve D_A depicts the pretax rate of return earned by capital in location A. The curve is downward sloping from left to right in recognition of diminishing productivity as additional capital is employed in A. The pretax rate of return for capital in jurisdiction B is shown by D_B. This curve is read from right to left and is also downward sloping from that perspective. If capital can move freely between the two jurisdictions, adjustments will be made until the rates of return, r_e, are equalized. This equilibrium is shown by point K_1 in Figure 18.6. Thus the optimal amount of capital in A is O_AK_1, and the amount in B is O_BK_1.



Now consider a 5 percent property tax imposed on capital in jurisdiction *A*. This tax reduces the after-tax rate of return on capital. The new rate of return is shown by D'_A. The impact of the tax is to cause capital to be moved from *A* to *B* until the rates of return are again equal. This occurs at K_2. Thus, the tax causes $K_2—K_1$ units of capital to be shifted from *A* to *B*. Note that the rate of return, r'_e, earned in both areas is lower than before the imposition of the tax. Thus, the property tax on mobile capital has caused a decline in wealth for the owners of capital in jurisdiction *B* as well as those in *A*.

In general, a property tax can be considered as a tax on certain inputs. As such, its effects are similar to those described in the preceding section. If substitution between inputs is possible, then, by increasing the relative costs of land and capital, the property tax tends to decrease the use of these inputs. But if the amounts of land and capital are fixed, production techniques remain unchanged and the property tax simply reduces the market value of the assets.

Key Concepts

- A tax imposed on one input causes a cost-minimizing manager to use more of other inputs and less of the taxed input.
- If property is fixed in location, a property tax will be capitalized. That is, the tax will reduce the market value of the asset.
- If capital is not fixed in location, it will tend to move from high-tax to low-tax jurisdictions until the rate of return is equalized. The rate of return in both jurisdictions will be lower than before the tax was imposed.

TAX PREFERENCES

Although taxes extract a substantial share of the earnings of most firms, it is possible for managers to reduce tax payments by taking advantage of various tax preferences that have been incorporated into the tax system. As used here, tax preferences refer to business activities, decisions, or conditions that are given preferential treatment under tax laws. Three examples embodied in certain income tax laws are considered in this section. First the provision that allows deduction of interest expenses is discussed. The second tax preference analyzed is an investment tax credit. Finally, the implications of using accelerated depreciation to compute tax liability are considered. In each case, the objective of the analysis is to show how these provisions affect managerial decisions.

Interest Deductions

Firms obtain money for investment using a combination of debt and equity finance. But these sources of funds are treated differently with respect to tax laws. Interest payments on debt are considered as expenses and can be deducted in computing tax liability. In contrast, the dividends paid to stockholders and funds kept by the firm as retained earnings are not deductible.

The preferential treatment of debt under the corporation income tax affects the optimal capital structure used to finance the firm. Consider a corporation that needs to obtain X dollars to pay for a new production facility. The investment is to be financed using a combination of debt and equity. Thus

$$D + E = X$$

where D is the amount of debt and E is the amount of equity.

If the cost of \$1 of debt is r_d and the cost of \$1 of equity finance is r_e, the total cost of financing the investment is given by $r_dD + r_eE$. But if the tax is imposed, the total cost becomes

$$(1 - t)r_dD + r_eE$$

where t is the rate of the corporation income tax. Note that the effective cost of debt is $(1 - t)r_d$ because deduction of the interest expense reduces the firm's tax liability by tr_d per dollar of debt.

Suppose that the objective of management is to minimize the total cost of financing the investment. One approach to the problem is to start with the assumption that only equity financing is to be used and then consider the effect on total cost of substituting dollars of debt for dollars of equity. Initially, it is assumed that r_d and r_e are not affected by the relative proportions of debt and equity.

Adding one more dollar of debt finance costs the firm $(1 - t)r_d$ while reducing equity finance by the same amount generates a saving of r_e. Hence, the net effect on total cost of financing the \$1 is

$$(1 - t)r_d - r_e \tag{18-1}$$

If costs of debt and equity finance are equal ($r_d = r_e$), then because of the tax deductibility of interest, the use of additional debt continually decreases the total cost. Generally, however, the cost of debt is less than the cost of equity because debt holders assume less risk than do those who have equity holdings. This is because interest on debt

must be paid before any money can be paid to a firm's stockholders. As compensation for the greater risk they bear, shareholders demand higher rates of return. If it is assumed that $r_d < r_e$, the advantage of debt finance is even greater.

Equation (18-1) has a somewhat surprising implication. In that additional dollars of debt always decrease the total cost of financing the $\$X$ of investments, costs will be minimized by using only debt finance. Clearly, this result is not consistent with actual practice. In fact, most firms have capital structures with more equity than debt. Other than utilities, relatively few large corporations have a capital structure with as much as 50 percent debt.

It should also be noted that the costs of debt and equity capital are probably not independent of the firm's capital structure. As relatively more debt is used, the level of risk also increases. This can be demonstrated by a simple example. Consider a corporation that earned $1,000,000 after payment to all inputs except capital. Suppose that the firm has an interest expense of $500,000. Thus, the amount of income left for dividends and retained earnings is $500,000. Note that earnings can decline substantially without impairing the firm's ability to meet its interest obligation. Of course, a decline in earnings would reduce the dollars left for dividends and retained earnings.

Now suppose that the firm's capital structure included proportionately more debt and that total interest expense was $950,000 instead of $500,000. If earnings are $1,000,000, the firm can still make its interest payments. But a relatively small decline in earnings can cause serious problems. If earnings drop by more than $50,000, the firm will be unable to meet its interest obligations. If this situation continues, the business may be forced into bankruptcy.

Investors require compensation for the increased risk associated with high debt/equity ratios. Thus, as more debt is used in the capital structure, the costs of both debt and equity finance increase. Let the higher costs of debt and equity (in comparison to the 100 percent equity case) be Δr_d and Δr_e, respectively. This increased cost is applied to the total dollars of debt (D) and the total dollars of equity (E). Thus the effect on total cost of using $1 more debt is given by

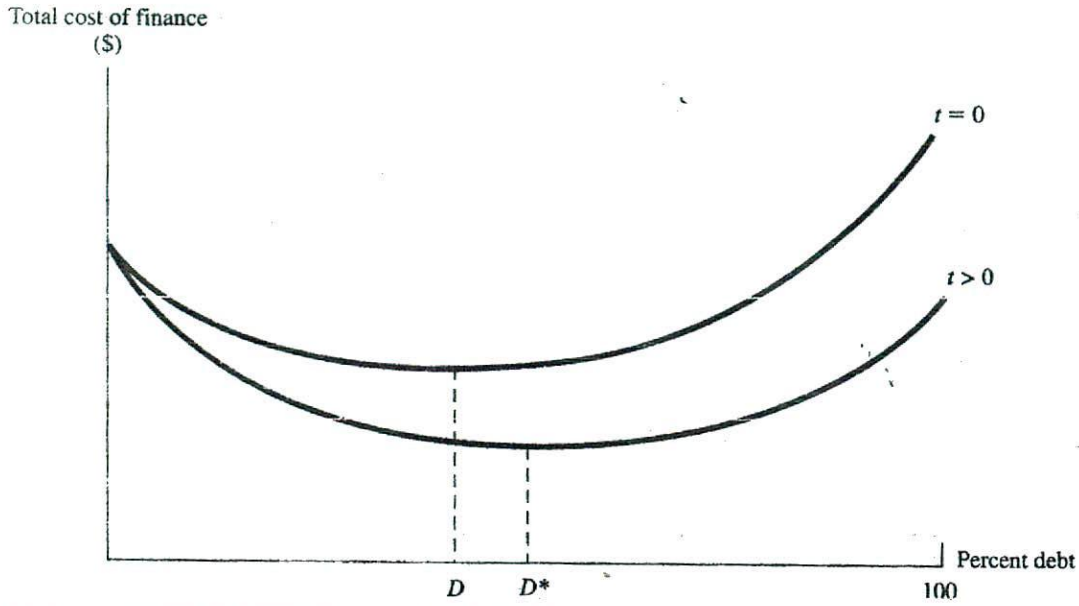
$$(1 - t)(r_d + \Delta r_d D) + (-r_e + \Delta r_e E)$$

which can be written as

$$[(1 - t)r_d - r_e] + [(1 - t)\,\Delta r_d D + \Delta r_e E] \qquad (18\text{-}2)$$

Equation (18-2) is easily interpreted. The first term in brackets is identical to equation (18-1) and represents the basic cost advantage of an additional dollar of debt finance. Note that for $r_d \leq r_e$, this term is always negative, indicating that total cost is reduced by using more debt. The second term in brackets is always positive and reflects the increased costs of capital associated with higher debt/equity ratios. As additional debt is utilized, $(1 - t)\,\Delta r_d D$ and $\Delta r_e E$ both become greater and thus the risk premium increases.

Equation (18-2) can be used to determine the optimal capital structure. Additional debt should be added until the basic cost advantage of debt is just offset by the risk premium required for a higher debt/equity ratio. Initially, the risk premium will be small and total costs will be reduced by using more debt. But beyond some point, adding debt increases the total cost. This result is shown graphically in Figure 18.7. With no taxes (i.e., $t = 0$), the curve indicates that financing costs are minimized when the capital structures includes D percent of debt.



Equation (18-2) shows how the tax preference for interest expenses affects the optimal capital structure. If $t = 0$, debt will be used only as long as the pretax cost advantage of debt ($r_d - r_e$) exceeds the risk premium. But for $t > 0$, by reducing the cost of debt relative to equity, the tax system provides an incentive for managers to use more debt. Hence, as shown by D^* in Figure 18.7, debt/equity ratios are likely to be higher when there is a tax preference. Note that the tax preference reduces the total cost of finance for all debt/equity ratios.

Tax-Exempt Fringe Benefits

Fringe benefits are compensation to workers in addition to their wages or salaries. Examples include paid vacations, medical and dental insurance, life insurance, and retirement programs. In the past 30 years, fringe benefits as a proportion of the total wage bill have increased, especially in union contracts. In 1960 employers of union workers paid 21 cents in fringe benefits for every dollar paid as wages. By 1998 that cost had nearly doubled, to 39 cents per wage dollar. The popularity of fringe benefits can be partially explained by tax preferences for such benefits.

In hiring workers, a manager must consider the total cost of the compensation package being offered. This package includes a salary plus various fringe benefits. In evaluating a job offer, workers may be willing to trade salary for additional benefits. However, the terms of trade can be affected by the tax treatment of fringe benefits.

Certain fringe benefits are exempt from the personal income tax. For example, medical insurance and retirement contributions can be provided by an employer and not considered as taxable income to the worker. Thus, because of the tax advantage, the worker may be better off taking an additional dollar in fringe benefits rather than in

salary. But the costs to the firm are the same. There is no tax advantage to the firm because both salary and fringe benefits can be deducted in computing the firm's income tax bill. Hence, from the cost perspective of the firm, a dollar of salary is equivalent to a dollar of fringe benefits.

The preferential treatment of fringe benefits may alter the compensation package offered by the firm. Because the after-tax value to the worker of an additional dollar of fringe benefits is greater than an extra dollar of salary, the firm can provide the same total compensation to the worker at a reduced cost by offering relatively more fringe benefits than if the tax advantage were not available.

Investment Tax Credits

Business investment is crucial to a market economy. Without sufficient investment there will be little economic growth, and existing facilities will soon become obsolete. Tax policy can be used to stimulate investment. One approach is the granting of tax preferences for investment. An investment tax credit is an example of such a preference. This credit allows businesses to reduce their corporation income tax liability by some fraction of the firm's investment spending during the year.[1]

In chapter 15 it was argued that firms should undertake only those investments that have positive net present value. Consider an investment that requires an initial outlay of C dollars and has a useful life of T years. If the discount rate is r and the after-tax revenue resulting from the investment in a given year is $(1 - t)R_i$, the net present value of the investment is

$$NPV = \sum_{i=1}^{T}\left[\frac{(1-t)R_i}{(1+r)^i}\right] - C$$

If $NPV > 0$, the investment is profitable for the firm.

Now suppose that the provisions of the corporation income tax are changed to allow a 10 percent tax credit for investments made during the tax year. This credit would allow the firm to subtract $0.10C$ from its income tax payment, meaning that the actual cost to the firm of the investment would be only $0.90C$. Hence, the net present value of the investment is now

$$NPV = \sum_{i=1}^{T}\left[\frac{(1-t)R_i}{(1+r)^i}\right] - 0.90C \tag{18-3}$$

Because costs are reduced, the net present value increases. Thus, one effect of the investment credit is to make profitable investment proposals more profitable. Another is to stimulate investment in projects for which net present value had been negative but that now are profitable because the net present value is positive. A third effect is to cause substitution of capital for other inputs in production. Because the effective after-tax cost of capital equipment is less than it would be without the credit, capital becomes relatively less expensive. Hence, costs can be reduced by using more capital and less of other inputs.

[1] The general tax credit provision of the Federal Corporation Income Tax was repealed by the Tax Reform Act of 1986. However, investment tax credits still exist at the state level and for certain types of equipment at the federal level.

Case Study

Solar and Conservation Credits: Incentive or Windfall?

Tax credits can be used to alter consumer purchasing decisions. During the late 1970s, there was a strong sentiment in Congress to conserve energy and promote the use of alternative energy sources. Legislation was enacted that allowed taxpayers a credit against their income taxes of 15 percent of the cost of home energy conservation expenditures to a maximum tax saving of $300. A credit was also allowed for the purchase of solar, wind, and other alternative energy systems. This credit was set at 40 percent of the cost of the system, with a maximum tax saving of $4,000.

The effect of the conservation and alternative energy system tax credits was to reduce the cost of such investments. By making them less expensive relative to use of conventional fuels such as oil and gas, the intent was to induce consumers to increase their conservation and alternative energy system expenditures.

However, one concern about the consumer tax credits is whether they significantly change consumer purchasing patterns. Those consumers who would have invested in conservation or alternative energy systems even without the tax credits receive a windfall benefit. In evaluating the credits, the windfall must be compared to the benefits to society of the additional investment that was induced by the tax credits.

Although the evidence is not definitive, the case for conservation tax credits appears weak. Survey information collected by the authors suggests that almost all conservation expenditures would have been made even without the tax credits. In contrast, credits for alternative energy investments seem to have played an important role in stimulating investment in this area. Only a small proportion of the survey respondents said they would have bought solar or wind energy systems if the credits had not been available.* ■

*H. C. Petersen, "Solar Versus Conservation Tax Credits," *Energy Journal* 6 (3) (1985): 129–135.

Accelerated Depreciation

In computing corporation income tax liability, depreciation expenses are deductible. Thus, when depreciation is taken into account, the net present value of an investment is given by the expression

$$NPV = \sum_{i=1}^{T}\left[\frac{(1-t)R_i + td_iC}{(1+r)^i}\right] - C$$

where d_i is the fraction of the total cost of the investment that can be depreciated in the ith year. Note that the depreciation allowance for any given year increases the net present value of the investment by $td_iC/(1+r)^i$. The reason is that the present value of the

firm's tax liability is reduced by that amount. The expression for net present value can also be written as

$$NPV = (1 - t)\sum_{i=1}^{T}\left[\frac{R_i}{(1 + r)^i}\right] + tC\sum_{i=1}^{T}\left[\frac{d_i}{(1 + r)^i}\right] - C \tag{18-4}$$

For given C and t, note that the net present value increases as $\sum_{i=1}^{T} [d_i/(1 + r)^i]$ increases. But $\sum_{i=1}^{T} d_i = 1$. Thus, the magnitude of $\sum_{i=1}^{T} [d_i/(1 + r)^i]$ depends on the size of the individual d_i. Because the depreciation benefit is discounted to the present, a method that allows larger write-offs in the early years of the depreciation period will result in greater net present value than if the depreciation rate is constant over the useful life of the investment. Thus, if a firm is allowed to depreciate its assets more rapidly, otherwise unprofitable investment proposals may now show a positive net present value. The result is that additional investment may occur.

The straight-line method is the simplest technique for computing depreciation allowances. For a depreciation time of T years, this approach would specify $d_i = 1/T$ for all i. For example, if $T = 5$ and the cost of an investment is \$100,000, the annual depreciation allowance is one-fifth of \$100,000, or \$20,000.

To stimulate additional investment, firms are sometimes allowed to use accelerated depreciation methods. The double-declining-balance technique is one example. Instead of deducting C/T in the first year, the firm is allowed to deduct $2/T$ of the total amount in the first year and $2/T$ of the remaining balance in subsequent years. For the example just given, 40 percent of the \$100,000 could be depreciated in the first year (\$40,000), 40 percent of the remaining \$60,000 in the second year (\$24,000), 40 percent of \$36,000 in the third year (\$14,400), and 40 percent of the remaining \$21,600 in the fourth year (\$8,640). The final \$12,960 of the original \$100,000 would be written off in the last year of the depreciation period.

Although the total amount of depreciation is \$100,000 under both schemes, the present value of the tax saving is greater with accelerated depreciation. If the tax rate is 50 percent and the discount rate is 10 percent, the present value of the tax saving for straight-line depreciation is

$$(0.50)\left[\frac{20{,}000}{(1.10)^1} + \frac{20{,}000}{(1.10)^2} + \frac{20{,}000}{(1.10)^3} + \frac{20{,}000}{(1.10)^4} + \frac{20{,}000}{(1.10)^5}\right] = \$37{,}900$$

while the amount using double-declining balance is

$$(0.50)\left[\frac{40{,}000}{(1.10)^1} + \frac{24{,}000}{(1.10)^2} + \frac{14{,}400}{(1.10)^3} + \frac{8{,}640}{(1.10)^4} + \frac{12{,}960}{(1.10)^5}\right] = \$40{,}515$$

Note that the net present value of the investment increases by \$2,615 under the double-declining-balance method. Thus the investment is more likely to have a positive net present value if accelerated depreciation is allowed for tax purposes.

Key Concepts

- Deductibility of interest payments on debt has the effect of increasing the amount of debt in a firm's optimal capital structure.
- Investment tax credits may stimulate investment by reducing the cost and hence increasing the net present value of an investment.
- By increasing the present value of tax savings, accelerated depreciation increases the net present value of an investment. Thus, the rate of investment will increase.

SUMMARY

By increasing the effective price, an excise tax can be used to decrease the demand for a good or service. However, the effect of an excise tax on the profit-maximizing price charged by a firm depends on conditions of supply and demand. The proportion of the tax that can be passed on to consumers is inversely related to the elasticity of demand. If demand is totally elastic, an excise tax must be absorbed by sellers. Conversely, the price increase that is caused by an excise tax is directly related to the elasticity of supply. If the supply curve is horizontal, consumers will pay the entire tax.

The effect of a tax on profit depends on the objectives of managers. If the goal of managers is to maximize profit, a tax on profit will not affect the optimal price and quantity in the short run. However, if the objective is revenue maximization subject to a profit constraint, a profit tax will cause a reduction in output and an increase in price in the short run.

Excise taxes may be imposed on inputs used by a firm. Such taxes signal to managers that the relative prices of the inputs have changed. If technology allows substitution, a tax on one input will cause other inputs to be substituted for the input that is taxed.

Property taxes reduce the rate of return earned on the taxed property. If the property is fixed in location, the property tax will be capitalized and the market value of the property will decline. But if the property can be moved from one location to another, firms will shift property from jurisdictions with high tax rates to those with lower rates. This adjustment will continue until the after-tax rates of return are equal in all jurisdictions.

Interest payments on debt can be deducted in computing income taxes. This preferential treatment of debt reduces its after-tax cost relative to equity. Hence, the tax system causes relatively more debt to be used in the corporation's capital structure. But higher debt/equity ratios mean greater risk for those who provide funds to the firm. Thus, the costs of debt and equity capital increase. The implication is that there is a cost-minimizing debt/equity ratio for the firm.

The popularity of fringe benefits can be partially explained by tax preferences. Because certain benefits are exempt from the personal income tax, a dollar of such benefit is worth more to the worker than a dollar in salary.

Investment tax credits are used to stimulate investment in capital goods. They allow the firm to reduce its income tax by some fraction of the cost of an investment. By increasing their net present value, some investments that would have been unprofitable without the credits may now be undertaken by the firm.

By treating depreciation as a deductible expense, the tax system increases the net present value of an investment by an amount equal to the present value of the tax saving. But methods of accelerated depreciation, such as the double-declining-balance approach, increase the present value of the tax saving. As a result, they stimulate investment demand and encourage the substitution of capital for other inputs.

Discussion Questions

18-1. An excise tax is imposed on a product for which there are few good substitutes. Who will pay the tax, the firm or the consumer? Explain.

18-2. If there is no minimum profit constraint, how does the revenue-maximizing firm select the optimal rate of output?

18-3. A good is produced using capital and labor. Suppose that a tax is imposed on capital. What is the relationship between the convexity (i.e., curvature) of the isoquant and the effect of the tax? Explain.

18-4. What does it mean to say that a tax has been capitalized?

18-5. A tax on capital used in jurisdiction *A* could reduce the rate of return on capital employed in jurisdiction *B* even though there is no tax on capital in *B*. Explain.

18-6. New legislation eliminates all taxes on profit. How would this change affect the optimal proportions of debt and equity in the capital structure of firms?

18-7. Why does the marginal cost of both debt and equity increase if relatively more debt is used in the firm's capital structure?

18-8. A firm operates in a country that imposes a highly progressive tax on profit. Interest paid on debt can be deducted from taxable income and there is an investment tax credit. How is the effect of the interest deduction related to the profitability of the firm? How is the effect of the investment tax credit related to the profitability of the firm?

18-9. For tax purposes, firms are allowed to use accelerated depreciation. In evaluating investment decisions, how is the effect of accelerated depreciation related to the discount rate used by the firm in decision making?

Problems

18-1. The market supply and demand functions for videotapes are given by

$$Q_S = 100 + 20P$$
$$Q_D = 300 - 5P$$

where Q is quantity and P is the price of tapes.

a. What are the equilibrium price and rate of output?

b. If an excise tax of \$2 per tape is imposed on the seller, what will be the new equilibrium price and rate of output? What proportion of the tax will be paid by purchasers of the tapes?

18-2. The market supply-and-demand functions for deluxe pizzas in a small town are given by:

Demand: $Q = 100 - 3.5P$

Supply: $Q = 15 + 1.5P$

a. Determine the equilibrium price and quantity.
b. If the city government levies a tax of $3.00 per pizza on the pizza parlor, determine the new equilibrium price and quantity of pizza.
c. What portion of the tax will be paid by the buyer and what portion will be paid by the seller?

18-3. In a small, isolated college community, there are 2,000 apartments. In the short run, the only use of these apartments is for rental to students. The college has a large summer school program that results in a 100 percent occupancy rate throughout the year. The demand for apartments by students is given by

$$Q_D = 4{,}000 - 2P$$

where Q_D is the number of apartments demanded and P is the average monthly rental rate.

Graffiti painted on walls and sidewalks is a serious problem in the community. The mayor thinks that college students are responsible and that they should pay for the cleanup. Because students occupy all of the town's apartments, she proposes a monthly tax of $25 on each apartment rental.

a. On an annual basis, how much revenue will the tax provide?
b. Evaluate the tax as a means of forcing students to pay for cleanup of their graffiti.

18-4. Fred Merkle and Co. is the monopoly supplier of old movies in a region. Over a year's time, the demand for movies is given by

$$P = 100 - Q$$

where P is the price of movies and Q is the quantity demanded per year. The firm's total cost (TC) and marginal cost (MC) functions are

$$TC = 800 + 20Q + Q^2$$

and

$$MC = 20 + 2Q$$

a. What are the profit-maximizing price and rate of output?
b. What are the revenue-maximizing price and rate of output?
c. If the objective is to maximize revenue with a constraint that total revenue be greater than or equal to total cost, what are the revenue-maximizing price and rate of output?
d. Repeat part (c), assuming that a 10 percent tax is imposed on profit. Is your answer consistent with the discussion of the profit tax in the chapter? Explain.

18-5. A firm manufactures a product using both capital and labor. A federal tax credit is available for capital expenditures. Use isoquant analysis (see Figure 18.5) to show how the optimal combination of inputs used to produce Q_0 units of output would be affected by the tax credit.

18-6. A parcel of land has a market value of $100,000 and is located in a county with no property tax. Assume that a permanent $500 per year property tax is imposed on the land. If the tax is fully capitalized, approximately what will be the market value of the land after the imposition of the tax? Assume that investors use a discount rate of 10 percent.

18-7. A firm's production is based on the equation $Q = 10K^{0.5}L^{0.5}$. The price of capital is \$5 and the price of labor is \$2. The marginal product of labor is $MP_L = 5K^{0.5}L^{-0.5}$ and the marginal product of capital is $MP_K = 5K^{-0.5}L^{0.5}$.
   a. To minimize costs, how much labor and how much capital should the firm employ to produce 100 units of output?
   b. If a tax of \$1 is imposed on each unit of labor, what would be the cost minimizing amounts of labor and capital to produce 100 units? What if a tax of \$1 is imposed on both labor and capital?

18-8. Managers of Quick Foto of Fargo, North Dakota, are considering the purchase of a new device for processing film. The cost of the machine is \$10,000; it has an expected useful life of 10 years and a salvage value of \$1,000. The firm's managers believe that the machine will increase Quick Foto's revenues by \$2,000 per year. Operating expenditures are projected to be \$500 per year. Assume that revenues and operating expenditures are incurred at the end of each year and that managers use a 14 percent discount rate in making investment decisions. The firm's marginal tax rate is 20 percent. What rate of investment tax credit would be necessary to cause managers to purchase the machine? Assume that the credit is received at the time of purchase.

18-9. A new cement truck would cost the Sure Stick Concrete Company \$100,000. The law allows the firm to use either straight-line or double-declining-balance depreciation for tax purposes. The truck can be depreciated over 5 years and the discount rate used by managers is 12 percent. The firm's marginal tax rate is 30 percent. What effect does the choice of a depreciation method have on the net present value of the investment in the truck?

Problems Requiring Calculus

18-10. A property tax of T dollars per year is placed on a piece of urban property. If the tax is fully capitalized,
   a. Write a general equation showing the change in the market value of the property as the amount of the tax changes.
   b. Write a general equation showing the change in the market value of the property as the discount rate used to capitalize the tax changes.

18-11. Acme Manufacturing produces a product using labor and capital as inputs. The firm's production function is given by

$$Q = 25K^{0.1}L^{0.9}$$

The price of labor is \$10 and the price of capital is \$20.
   a. If the product is to be manufactured at minimum cost, for any rate of output, how much labor should be used for each unit of capital employed?
   b. Suppose a 10 percent tax is imposed on capital. How much labor should be used for each unit of capital?
   c. Starting from the initial prices, if a 10 percent tax is placed on each input, how much labor should be used for each unit of capital employed?

CHAPTER 19
Antitrust and Regulation

PREVIEW

Because competition benefits society by reducing prices and improving the efficiency of resource allocation, a top priority for government action should be the implementation of policies designed to enhance competition. In the United States, the primary public policy approach to increasing competition is through the use of antitrust laws. These statutes give enforcement agencies the power to alter existing or proposed market structures and to impose penalties for certain types of business conduct determined to be anticompetitive.

To avoid litigation, it is important that managers understand the scope and limits of antitrust law. Thus, the first section of this chapter considers antitrust activity in the United States. The objective is to make the reader aware of business activities that may violate antitrust statutes. The discussion begins with a review of basic U.S. antitrust laws and enforcement procedures. Next is a consideration of antitrust law as it applies to monopolization, mergers, collusion, and price discrimination. Finally, remedies and penalties used in antitrust enforcement are discussed.

Antitrust laws deal with industry structure and conduct. In contrast, the regulatory approach to public policy focuses on industry performance. It is based on the presumption that there may be circumstances where competition is not possible or is not desirable. Thus, to prevent adverse consequences from the lack of competition, government may regulate firms so that they perform in a socially acceptable manner. Regulation may take the form of requiring approval for price changes, limiting entry or exit, or prescribing standards that a product or service must meet. The second major section of the chapter examines public utility regulation in the United States. The first topic considered is the need for public utility regulation. Following that is a brief overview of regulatory procedures.

Price controls are another tool that can be used by government to control business activity. Generally, they are used to combat inflationary pressures and expectations. The last section of this chapter evaluates the impact of price controls. Of particular importance is the differential effect of controls in competitive and concentrated industries.

ANTITRUST POLICY

Most industries do not meet the criteria for perfect competition as discussed in chapter 9. Moreover, the complexities of markets and the politics of decision making make it unlikely that any reasonable set of public policies could generate the conditions for perfect competition. Thus, public policy must be content with a more limited objective. It has been suggested that a realistic goal for policymakers is workable competition. It may not be necessary for all the requirements of perfect competition to be met to achieve results that approximate those of competition. Markets may fail to meet one or more of the criteria and still be "workably" competitive.

The achievement of workable competition is the goal of antitrust activity. The antitrust approach acknowledges that imperfections exist in many markets but is directed toward narrowing the gap between actual conditions and the competitive ideal. The philosophy underlying antitrust enforcement is that by modifying the structure of markets and conduct of participants in markets, performance can be improved without direct government involvement in the daily decision making of managers. Prevention of a merger between two large corporations is an example of a structural modification. A fine assessed for fixing prices illustrates antitrust policy intended to alter business conduct.

U.S. Antitrust Laws

The Sherman Act and the Clayton Act (as amended) represent the primary legal basis for antitrust activity in the United States. The most important provisions of these two statutes are discussed here.

Sherman Act The Sherman Act was enacted in 1890 and has remained basically unchanged for over 100 years. It has two main sections:

> *Section 1.* Every contract, combination in the form of a trust or otherwise, or conspiracy, in restraint of trade or commerce among the several states, or with foreign nations, is hereby declared to be illegal. Every person who shall make any such contract or engage in any such combination or conspiracy shall be deemed guilty. . . .
>
> *Section 2.* Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several states, or with foreign nations, shall be deemed guilty. . . .

These two sections focus on different types of undesirable business behavior. Section 1 is intended to prohibit firms from conspiring to initiate and maintain practices not in the public interest. For example, an agreement among managers to fix prices would violate Section 1. The Sherman Act, Section 2, is designed to reduce market dominance. Firms that aggressively act to gain control of their markets may be in violation of Section 2. For example, in 1911 Standard Oil of New Jersey was found guilty of illegal monopolization of the market for refined oil. As a result, the firm was split into 30 separate companies. Most of the large oil companies in the United States today are the result of the Standard Oil dissolution.

Clayton Act The Clayton Act was enacted as a supplement to the Sherman Act. The intent of Congress was to provide legislation to prevent firms from obtaining monopoly power and also to specify specific business practices that are prohibited. The most important provisions of the act are contained in Sections 2 and 7.

> *Section 2.* It shall be unlawful for any person engaged in commerce, to discriminate in price between different purchasers of commodities . . . where the effect of such discrimination may be to substantially lessen competition or tend to create a monopoly in any line of commerce. . . .
>
> *Section 7.* No corporation engaged in commerce shall acquire, directly or indirectly, the whole or any part of the stock or other share capital of another corporation engaged also in commerce where the effect of such acquisition may be to substantially lessen competition between the corporation whose stock is so acquired and the corporation making the acquisition or to restrain such commerce in any section or community or tend to create a monopoly of any line of commerce.

These two sections each deal with a specific type of business practice. Section 2 is directed against certain types of price discrimination. Setting prices below cost to eliminate competition in a market is an example. Section 7 of the Clayton Act imposes restrictions on merger activity. A merger between General Motors and Ford could (and probably

would) be prevented based on Section 7. Notice that the language of the Clayton Act is not absolute. That is, price discrimination and mergers are prohibited only if they tend to "substantially lessen competition or create a monopoly in any line of commerce."

Two important amendments have modified the original Clayton Act. In 1936, the Robinson–Patmen Act was passed to broaden the Section 2 provisions against price discrimination. The new act was aimed at large retailers who could undersell their smaller competitors because they could buy merchandise at lower prices from manufacturers and wholesalers.

In 1950 the Celler–Kefauver Amendment was enacted to supersede the provisions of Section 7 of the Clayton Act. The original Section 7 focused on competition "between the corporation ... acquired and the corporation making the acquisition." This wording caused the courts to ignore the broader issue of a general lessening of competition. The result was that enforcement agencies found it difficult to prevent vertical and conglomerate mergers under the Clayton Act. Congress responded to this problem by passing the Celler–Kefauver Amendment, which amended Section 7 to read:

> That no corporation engaged in commerce shall acquire, directly or indirectly, the whole or any part of the stock or other share capital and no corporation subject to the jurisdiction of the Federal Trade Commission shall acquire the whole or any part of the assets of another corporation engaged also in commerce, where in any line of commerce in any section of the country, the effect of such acquisition may be substantially to lessen competition, or to tend to create a monopoly.

Note that the amendment reduced the emphasis on reduced competition between the merging firms and stressed the idea that the demonstration of a lessening of competition "in any line of commerce in any section of the country" could be used to prevent a merger. The effect of the change was to make vertical and conglomerate mergers subject to antitrust action.

Antitrust Enforcement

Antitrust proceedings are initiated in four ways. First the Antitrust Division of the Department of Justice may file a suit. If the suit is continued to the point of formal litigation, it is first heard in a federal district court. If either party wishes to contest the decision of the district court, the matter is taken to a circuit court of appeals and, if the justices are willing to hear the case, to the Supreme Court. The Antitrust Division's responsibility is limited to initiating and prosecuting a case. The courts must determine guilt and penalties.

The second path of antitrust enforcement is through the Federal Trade Commission. When the commission staff decides to issue a formal complaint and the matter is contested by the defendant, an initial hearing is held before an administrative law judge who is a part of the FTC. If the judge decides for the defendant, the matter is dropped. However, if the decision is to uphold the complaint, the case can be appealed to the five FTC commissioners. If their decision is again against the defendant, the matter can be appealed to the federal courts.

A third enforcement procedure involves state antitrust legislation. Most states have their own antitrust statutes. Typically, complaints are prosecuted by the state attorney general's office and decided by state courts. Appeals from decisions by the state supreme court can be taken to the federal court system.

The fourth method for dealing with alleged antitrust violations is litigation by private parties. Individuals or firms may file suits in the federal district courts. For example, a firm that believes it has been overcharged because its suppliers have fixed prices could sue under Section 1 of the Sherman Act. Appeals are heard by a circuit court of appeals and, ultimately, the U.S. Supreme Court. Private suits represent over 90 percent of all antitrust actions.

Rule of Reason versus Per Se Offenses

The standard of proof required for conviction in antitrust prosecutions differs with the nature of the alleged violation. Sometimes, although an apparent antitrust violation may have occurred, it is not clear that there has been a net injury to society. Such cases are decided under a *rule-of-reason standard.* In rule-of-reason proceedings, successful prosecution requires not only the demonstration that the act has been committed, but also that society will be better off by prohibiting, modifying, or punishing the act. In contrast, certain activities are judged illegal without the requirement that the specific antisocial effects be shown. These acts are referred to as *per se offenses.*

The per se and rule-of-reason standards represent different points along a continuum. They differ in the volume and detail of evidence required for a successful prosecution. Rule-of-reason cases require extensive evidence proving that an act has been committed and demonstrating the damage that has been caused. Per se cases only require proof that the offense has been committed. In a sense, per se violations can be thought of as being judged as if the nature of the offense automatically dictates that the social costs of the act are clearly greater than any possible benefits. Thus, the only real issue to be decided is the remedy or penalty in the case.

Not all antitrust violations fit into the tidy categories of being per se or rule-of-reason offenses. Still, there are some examples that can be cited as illustrations of each. Generally accepted as per se offenses are agreements to fix prices, divide markets between sellers, and to restrict or pool output. Activities evaluated under the rule-of-reason standard are mergers and monopolization of a market by a large firm or firms. As a rough guide, violations involving business conduct for which there is no strong justification are decided on a per se basis. In contrast, cases involving the structure of an industry usually are judged using the rule-of-reason approach.

Key Concepts

- Section 1 of the Sherman Act prohibits unfair business practices such as price fixing. Section 2 deals with monopolization.
- Section 2 of the Clayton Act (as amended by the Robinson–Patman Act) limits price discrimination.
- Section 7 of the Clayton Act (as amended by the Celler–Kefauver Act) is used to prohibit mergers that may result in a substantial lessening of competition.
- Antitrust actions can be initiated by the Department of Justice, the Federal Trade Commission, state officials, or private parties.
- Conviction of rule-of-reason offenses requires proof that the act has been committed and that social costs exceed benefits. Per se offenses require only a demonstration that the actions have occurred.

Monopoly

Section 2 of the Sherman Act prohibits monopolizing, attempting to monopolize, and conspiring to monopolize. However, it is important to note that the economic and legal definitions of monopoly differ. In the study of economics, monopoly is defined as a single seller. The legal interpretation is much less restrictive. As used in antitrust proceedings, firms are viewed as having monopoly power if they have a high degree of control over the price of a good or service.

Since its enactment, the most controversial point with respect to Section 2 has been the standard of proof required for successful prosecution. The issue has been whether the Sherman Act made monopoly power in and of itself an offense or whether the showing of industry dominance had to be accompanied by evidence of illegal practices. During the past 100 years, the Supreme Court has accepted both points of view. The position taken by the Court at any given time has had an overwhelming effect on the use of Section 2. When proof of illegal acts to achieve monopoly power has been required, there have been few Section 2 convictions. For example, between 1920 and 1945, the Court seemed to require some evidence of illegal activity. During that period there were very few successful Section 2 cases. When the Court has held that the government only had to show the existence of monopoly power, Section 2 cases have been more frequent and more successful.

The current interpretation of Section 2 represents a compromise between the two extreme positions. The mere existence of monopoly power is not sufficient for successful prosecution under Section 2. However, the government no longer is required to show that a firm has engaged in acts that, considered by themselves, represent antitrust violations. Instead, prosecutors can focus on patterns of business behavior that have the net effect of allowing a firm to gain and maintain a monopoly position. Under this interpretation, practices that would be legal when considered in isolation or when used by smaller firms may be grounds for conviction if used by a firm judged to have dominance in an industry. That is, the standard used by the courts to judge large firms is more rigorous than that applied to smaller firms. For example, aggressive price cutting in markets with competition, coupled with price increases in other markets, probably would be ignored by antitrust officials if practiced by a small firm. But the same actions by a firm with a large market share could result in a suit for illegal monopolization.[1]

It should also be noted that the Supreme Court is unlikely to receive many new Section 2 cases in the future. Over the last 10 years, the Justice Department and the Federal Trade Commission have essentially abandoned initiating cases requesting dissolution or divestiture of dominant firms. Their reluctance results from the substantial time and dollar commitment required to prosecute such cases, uncertainty as to their outcomes, and a perception that such cases are seldom necessary to protect consumers in today's global economy.

On those rare occasions when Section 2 cases are initiated, three conditions should be satisfied. First, substantial monopoly power must exist and must have been exercised. High market shares and findings of above-average profits are necessary but not sufficient conditions to satisfy this criterion. Second, possession of market power should be the result of more than just superior products sold by the firm or outstanding business

[1]For example, see *United States v. Grinnell Corp.*, 384 U.S. 563 (1966).

ability displayed by its management. There should be evidence that the firm has used its market power to suppress competition. Even then, a careful evaluation of the firm's practices is required to determine that they have no efficiency-enhancing effects on the economy. Third, a case should not be initiated unless there is some identifiable remedy that will provide net benefits to society. This is crucial because once an enterprise has been restructured, any efficiency benefits may be lost forever. If the social costs associated with the status quo are small or are likely to be transitory, Section 2 remedies may not increase social welfare.

Most students of antitrust policy would agree that Section 2 of the Sherman Act has probably been a modest deterrent to high concentration in an industry. However, the main impact may come not from the results of litigation, but as firms modify their plans to avoid the costs and uncertainties of possible prosecution. It has been suggested that the ghost of Senator Sherman sits in every corporate boardroom.

Case Study

The Great Antitrust Doubleheader

In January 1969, the Justice Department brought suit against IBM under Section 2 of the Sherman Act for illegal monopolization of the general-purpose computer market. The government proposed that the firm be split into several competing companies. It took 6 years for the case to come to court. One reason for the delay was the discovery process whereby the government was required to make available over 25 million pages of documents to IBM, and the firm provided over 60 million pages of documents to be used by the government in case preparation. The trial at the District Court level lasted 6 more years and generated 300,000 pages of testimony and exhibits.

In 1972, the Justice Department initiated litigation to force the American Telephone and Telegraph Company to sell off its manufacturing arm, Western Electric, and its local operating companies. The government charged that AT&T had used unfair practices to eliminate competition in the markets for telephone equipment and long-distance telephone service. The case has been referred to as the most important antitrust proceeding of all time because it sought the breakup of the largest privately owned enterprise that the world has ever known. At the time the suit was filed, AT&T employed several hundred thousand people and had revenues greater than the gross domestic product of all but 12 nations.

On January 8, 1982, the Justice Department announced resolution of both suits. The IBM suit was dropped because of changing conditions that had reduced the firm's market share, and hence diminished the government's chances of obtaining a conviction. The AT&T suit was settled by a consent decree that required the firm to divest itself of all its local operating companies. These operating companies, such as Pacific Telephone and Telegraph, represented about two-thirds of AT&T's total assets. In return, the firm was allowed to keep Western Electric and its long-distance operations and to become an active competitor in selling computers, an activity that had been prohibited by an earlier consent decree.

The settlement of the two cases on the same day probably was not a coincidence. While dropping one suit, the government could claim victory in another, thus avoiding political criticism. Also, although the IBM suit was terminated, a powerful new competitor, AT&T, was freed to compete with IBM in the computer industry.

The objectives of these antitrust actions have been only partially realized. Competition in many telecommunications markets has increased, and consumers have a wider range of choice than before divestiture. However, AT&T has failed in its attempts to be an effective competitor in the computer industry. ■

Merger

It is more difficult to constrain existing market power than to halt monopolization in its formative stages. Because industry dominance has frequently been achieved by acquiring other firms, merger policy is an important tool for antitrust action.

Evaluation of Mergers Mergers between large firms create the potential for abuse of the market power obtained by the combined firm. This abuse may take the form of higher prices, actions to deter entry or to eliminate competitors, or exercise of monopsony power to obtain price reductions from suppliers.

The primary argument in support of mergers is the possibility of efficiency gains. These efficiency effects may result from several different factors. The merged firm may be able to reduce its cost of production and distribution by realizing economies of scale. Inefficient techniques may be abandoned as firms gain access to the patent rights and technical expertise of their new partners. Mergers may also concentrate assets under the direction of superior managers who have the ability to operate the firm more efficiently.

Although there may be other considerations, the basic decision to allow or prohibit a merger should rest on an evaluation of the costs of increased market power versus the social benefits of improved efficiency. Williamson has suggested a simple model for quantifying this trade-off.[2] Consider the demand and cost curves shown in Figure 19.1. Suppose that a merger shifts the newly combined firm's cost curve down from AC_1 to AC_2, while providing market power that is exercised by increasing prices from P_1 to P_2. The welfare trade-off is shown in the figure. The crosshatched area A_1 represents the resource saving associated with producing Q_2 units at the reduced average cost, AC_2. The crosshatched area A_2 is the deadweight loss stemming from increasing the price from P_1 to P_2. It represents the loss in consumer surplus as the higher price reduces the quantity demanded from Q_1 to Q_2. The merger can be evaluated by comparing the cost saving and the deadweight loss.[3] If the deadweight loss is greater than the efficiency gain, the merger should not be allowed. But if the resource saving more than offsets the deadweight loss, there may be a net benefit to society from the merger.

---

[2] O. Williamson, "Economies as an Antitrust Defense: The Welfare Tradeoffs," *American Economic Review* 58 (March 1968): 18–36.

[3] Consumer surplus and deadweight loss are discussed on pages 328–329.

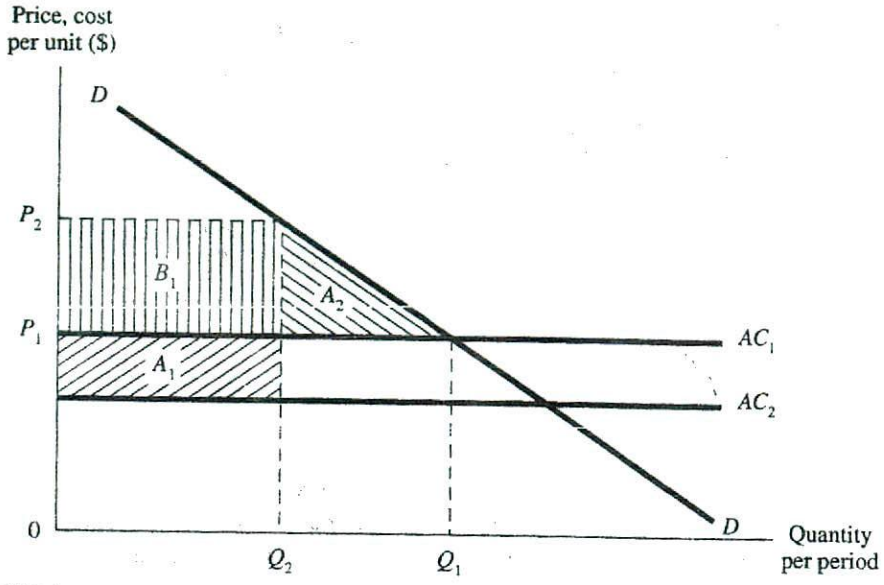


FIGURE 19.1 Evaluation of Horizontal Mergers

However, income distribution effects may also be important in evaluating a merger. Prior to the merger, the rectangle B_1 was part of the surplus value received by consumers. After the merger, this area represents economic profit earned by the combined firm. This transfer of value from consumers to the owners of the firm may be viewed as a reason for prohibiting the combination.

Merger Guidelines A merger between large corporations is a costly undertaking. Considerable expense is involved in finding a proper merger partner, structuring the merger proposal, communicating information to shareholders, and integrating the acquired firm. This cost can be greatly increased if the merger is challenged by the government. The actual expense of litigation can be substantial, but the delay and uncertainty involved are even more costly.

In 1984, the Justice Department announced guidelines to be used in deciding whether to challenge a merger. These guidelines are based on overall industry concentration as measured by the Herfindahl Index. This index is computed by summing the squared market shares of all the firms in an industry. That is,

$$HI = \sum_{i=1}^{m} X_i^2$$

where m is the number of firms and X_i is the percentage market share of the ith firm. By squaring the market shares, the Herfindahl Index gives greater weight to large firms in an industry. For example, an industry with 10 firms each with a 10 percent market share would have a Herfindahl Index of 1,000. The index for an industry that has a dominant firm with a 50 percent market share and five smaller firms each with 10 percent shares would be 3,000.

TABLE 19.1 Likelihood of a Justice Department Challenge Based on 1984 Guidelines

	Postmerger Herfindahl Index		
Increase in the Herfindahl Index	*Less than 1,000*	*1,000–1,800*	*Greater than 1,800*
0	Will not*	Unlikely	Unlikely
50	Will not*	Unlikely	Likely
100	Will not*	Likely	Will*

*Except in extraordinary cases.

The first step in applying the guidelines involves an examination of the Herfindahl Indexes for the industry. If the postmerger index is less than 1,000, the merger ordinarily will not be challenged. Mergers where the postmerger index is between 1,000 and 1,800 and where the increase in the Herfindahl Index as a result of the merger is less than 100 points are unlikely to be challenged. But if the postmerger index is between 1,000 and 1,800 and the increase is greater than 100 points, or if the postmerger index is greater than 1,800 and the increase is greater than 50 points, the Justice Department is likely to challenge a horizontal merger. These rules are summarized in Table 19.1.

Although industry concentration as measured by the Herfindahl Index is the starting point in using the guidelines, other factors are also considered by the Justice Department in deciding whether to challenge a horizontal merger. These are incorporated into the decision process in recognition that market shares are not the only data that affect the competitive effects of mergers. Other factors include ease of entry into the industry, the financial condition of the firm being acquired, the impact of foreign competition, and possibly efficiency gains from a merger.

Example Horizontal Mergers: To Challenge or Not to Challenge?

An industry consists of eight firms with the following market shares:

Firm	*Market Share (%)*
1	30
2	10
3	10
4	10
5	10
6	10
7	10
8	10

Firm 1 announces that it intends to acquire firm 2. Will this merger be challenged by the Justice Department? Would a merger between firm 2 and firm 3 be challenged?

Solution Prior to any mergers, the Herfindahl Index for the industry is $30^2 + 7(10^2)$, or 1,600. After a merger between firm 1 and firm 2, the index would be $40^2 + 6(10^2)$, or 2,200. Because the Herfindahl Index is greater than 1,800 and the increase is more than 100, the Justice Department would challenge the merger.

If firm 2 and firm 3 were to combine, the postmerger Herfindahl Index would be $30^2 + 20^2 + 5(10^2)$, or 1,800. Because the index increased by more than 100 points, from 1,600 to 1,800, the merger is likely to be challenged. However, in making a final decision, Justice Department officials would consider other factors, such as efficiency gains that might result from the merger.

Present Merger Policy Horizontal mergers between large direct competitors may be challenged by the government. On most occasions, the courts have supported the government and prevented such mergers. The outcomes of proposed vertical and conglomerate mergers are less certain. During the 1960s and early 1970s, the government frequently was successful in challenging such mergers if they involved large firms. But during the late 1970s and 1980s, antitrust enforcement agencies have shown less interest in preventing vertical and conglomerate mergers. At present, unless it appears that such mergers would substantially increase horizontal market power, they are unlikely to be challenged.

Recently, federal antitrust officials, instead of simply blocking certain large horizontal mergers, have allowed the mergers to proceed after eliminating any anticompetitive effects—an approach that has been referred to as "fix it first." For example, as a condition for not opposing a merger between two large oil companies, the federal government might require the acquiring firm to sell off those operations where the firms were direct competitors with substantial market shares. This divestiture could involve refineries, pipelines, and/or service stations.

Key Concepts

- Practices considered acceptable when practiced by small firms may represent antitrust violations if used by a large firm with considerable market power.
- One method of evaluating a merger is to compare efficiency gains with the social costs of increased market power. The income distribution effects may also be important.
- The Herfindahl Index is computed by summing the squares of the market shares of all firms in the industry. It takes into account the size distribution of firms.
- Mergers are likely to be challenged by the federal government in industries if the postmerger Herfindahl Index is greater than 1,800 and where the change in the index is greater than 100.
- Historically, the government has been successful in preventing horizontal mergers between large direct competitors. Recently, antitrust officials have used a "fix it first" policy with respect to many such mergers.

Collusion

Another important goal of antitrust activity is to prevent collusion. Firms in oligopolistic markets have an incentive to collude. But the actual decision is based on a benefit–cost calculation by the firms involved. An important form of collusion is price fixing. The potential benefit of price fixing is obvious—increased profit. The costs fall into four interrelated categories. First, there is a cost associated with setting and changing industry price structures. Each firm in a cartel is in a different position, has different expectations, and has varying economic power in the industry. Hence, it may be difficult to reach agreement. A second cost of collusion is that imposed by the inevitable cheating

that will occur. Although profits can be increased by collusion, one firm can earn even greater profit by cutting its price slightly below the agreed level. Unless there is a mechanism for detecting and punishing cheaters, the cartel will soon fail as members cut prices to preserve their market shares.

A third cost of cartelization is that nonprice competition may replace price competition. For example, members of a cartel may engage in expensive advertising as a substitute for active price competition. Legal penalties imposed on convicted colluders are the fourth cost of price fixing. These may take the form of fines, prison sentences for executives, damage awards to private parties, or court orders to alter certain business practices.

The objective of antitrust policy is to reduce the net benefits of collusive activities. As the likelihood of conviction and the magnitude of penalties increase, the expected cost of violating antitrust laws increases. Also, if firms perceive that they are being actively scrutinized by antitrust authorities and that penalties from conviction will be severe, they will tend to adopt less easily detectable and less-effective methods of collusion.

The courts have consistently taken a hard line against collusion. The precedent-setting case (*United States* v. *Trenton Potteries Co. et al.*) involved 23 manufacturers of bathroom fixtures who had conspired to fix prices. Through their trade association, the manufacturers published standardized price lists, met to consider prices, and pressured one another to sell only at list prices. When the association was brought to trial, it claimed that the agreement had not injured the public. The trial record supported this position, indicating that fixtures were often sold below the established prices. But the Supreme Court rejected the request for a rule-of-reason interpretation of price fixing. The justices argued that

> the reasonable price fixed today may through economic and business changes become the unreasonable price of tomorrow. . . . Agreements which create such potential power may well be held to be in themselves unreasonable or unlawful restraints, without the necessity of minute inquiry whether a particular price is reasonable or unreasonable.[4]

The strong per se condemnation of price fixing has consistently been reaffirmed by the courts. Agreements to fix prices are a violation of the Sherman Act without regard to their effect. The prohibition applies not only to fixing minimum prices but also maximum prices and price differentials. Firms are simply not allowed to act in concert in determining prices.

Illegal price fixing costs billions of dollars each year in higher prices. Although explicit collusion has not been eliminated by antitrust efforts, the fact that it is judged by the courts as a per se violation of the law has had a significant impact in reducing the most effective forms of price fixing, group boycotts, and market allocation. Conspirators have been forced to abandon overt methods and to settle for less easily detectable and less-efficient methods of collusion.

---

[4] *United States* v. *Trenton Potteries Co. et al.*, 273 U.S. 392 (1927).

Case Study
The Electric Machinery Conspiracy

In the early 1960s, 29 corporations were successfully prosecuted for fixing prices of electrical equipment such as transformers, generators, and switchgear. The indictments alleged two primary types of conspiracies. For sales involving open bids, the firms simply met to fix the prices that would be charged for different types of equipment. It was the sealed bid sales to the government that made the case intriguing. The intent of the conspiracy was to raise prices while allowing firms to maintain a predetermined market share. For example, GE was allocated 42 percent of the market for switchgear, Westinghouse 38 percent, Allis-Chalmers 11 percent, and ITE Circuit Breakers 9 percent. Market shares were maintained by rotating bids so that each firm became the low bidder the requisite percentage of the time. This was accomplished by changing the order of bids about every two weeks or "with the phases of the moon."

To avoid detection, the conspirators engaged in elaborate precautions. Only first names were used and mail was always sent in plain envelopes to the homes of the executives. Calls were made from pay phones. Each firm was referred to by a code number. Expense-account vouchers were disguised by making them out for cities that were approximately as far from the firm's offices as the actual location of the conspiratorial gathering. The meetings themselves were often held in out-of-the-way places—a favorite was Dirty Helen's Bar in Milwaukee. Above all, company lawyers were never told anything.

As part of its investigation of the conspiracy, the Justice Department subpoenaed the records of the ITE Circuit Breaker Company. Nye Spencer, an employee of ITE, served as the scribe for the switchgear conspiracy. Spencer had kept detailed records of the conspiracy meetings to assist him in training his assistant. Confronted with the request for information, he turned over all his files to the government. Larger cracks in the dike appeared as lower-level executives cooperated with the investigation rather than implicate themselves further. Thus, the government was able to assemble a strong case that the conspirators decided not to contest. Following sessions of plea bargaining, an agreement was reached whereby the firms were allowed to enter reduced pleas to some of the indictments in return for pleading guilty to others.

In the penalty phase of the case, the judge sent seven defendants to jail for 30 days and granted suspended sentences to 20 others. Never before had business executives been sent to jail as a result of Sherman Act violations. The firms were also required to pay nearly $2 million in fines. However, the greatest cost to the corporations involved was 1,900 private suits, resulting in damage awards of over $400 million. ■

Price Discrimination

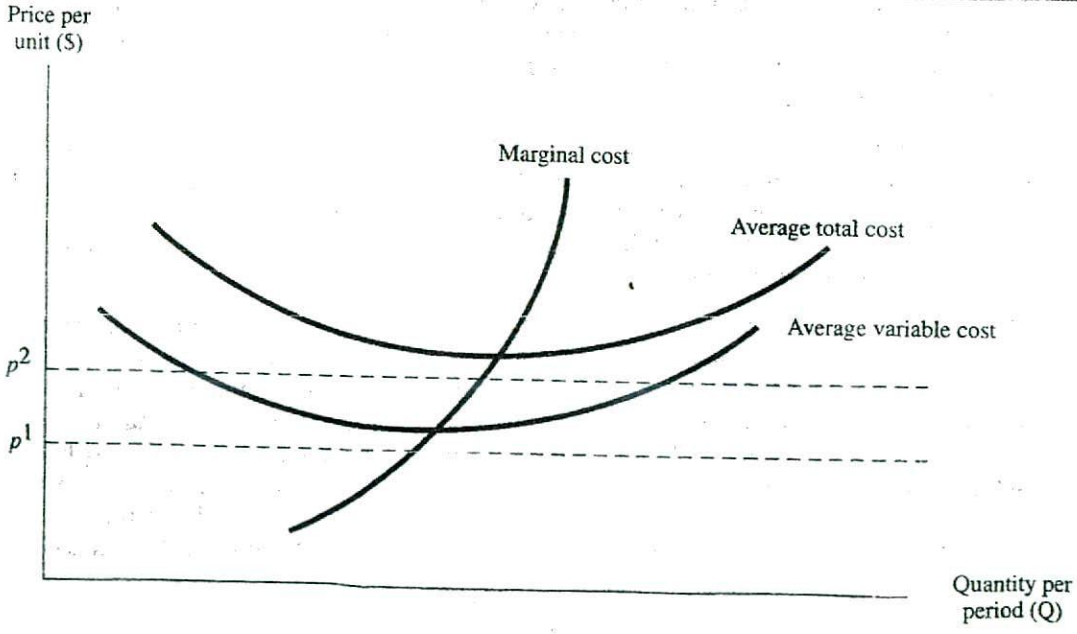
Price discrimination can be used to discourage entry and weaken existing competitors. Predatory pricing is an important example. Consider the case of a large firm selling a single product in a number of distinct geographical markets. In some of the markets the

firm has a monopoly, whereas in others it faces competition from smaller rivals operating in only a single area. The large firm can cut prices below cost in competitive markets and subsidize its losses from monopoly profits earned where the firm is the only seller. The smaller firms in the competitive markets may be forced from the market or into a merger if the dominant firm keeps its prices down. When the smaller firms have either gone out of business or been acquired, prices then are increased to the monopoly level.

Potential entrants into a market can also be deterred by the threat of predatory pricing. If the existing firm can create a credible threat that it will cut prices below cost if a new firm enters the markets, all but the largest potential competitors may be discouraged. Used in this manner, predatory pricing may be a useful tactic to prevent entry in an industry where natural barriers are not effective.

The courts have not been entirely consistent in defining what constitutes predatory pricing. However, recent court decisions have started from the premise that predation has occurred if price, p^1, is a set below average variable cost, as shown in Figure 19.2. From the perspective of economic theory, this position is defensible. In the short run, average variable cost represents the threshold between continuing to produce and shutting down an operation. If a price is below average variable cost, entrepreneurs can cut their losses by ceasing to produce. However, if price exceeds average variable cost, it is better to continue producing even if price is below average total cost, as shown by p^2 in Figure 19.2. The reason is that the excess can be used to pay a portion of the fixed costs. Thus, a price that is above average variable cost but below the firm's average total cost can be justified. In contrast, a firm that sets its price below average variable cost and continues to produce can be logically viewed as having other motives, such as the elimination of current or potential competitors.

FIGURE 19.2 Predatory Pricing



Price discrimination suits may involve considerable disagreement as to whether a firm has used predatory pricing. In some cases, an important issue has been whether there is evidence of intent to exclude competitors. But if the court determines that predation has occurred, the defendant will usually be found guilty of an antitrust violation.

Remedies and Penalties

Alleged antitrust violations are resolved in a number of different ways. The most common remedies and penalties are discussed here.

Consent Decrees Most antitrust actions are resolved by consent decrees. These are agreements worked out between the government's attorneys and those of the defendant. Usually, they specify certain activities that the firm must or must not do. In return, the government agrees not to prosecute. In accepting a consent decree the firm, in essence, says: "We didn't do it, but we won't do it again." An important advantage of consent decrees is that they cannot be used as evidence of guilt in other proceedings, such as a private antitrust suit.

Dissolution and Divestiture In monopoly and merger actions, the court may use dissolution or divestiture as a remedy. In 1911 the Supreme Court split Standard Oil of New Jersey into 30 smaller firms. This was an example of dissolution, whereby the firm loses its identity. In contrast, a divestiture order requires the firm to sell certain of its assets, but the firm retains its identity. In approving a 1984 merger between Socal and Gulf Oil, the Department of Justice required Socal to sell off several thousand Gulf retail gas stations. This was an example of divestiture.

Injunctions In ruling against a defendant, the court may issue an injunction that prohibits or compels certain actions on the part of the firm. For example, as a result of a price-fixing suit, a trade association may be prohibited from the collection and dissemination of information that was used to fix prices.

Fines Firms convicted of Sherman Act violations may be fined up to $1,000,000 per violation. Officers of the firms may receive fines to a maximum of $100,000. However, in some price-fixing cases, each day is considered a separate offense. In this way, the potential fines can be much greater.

Prison Sentences A 1974 amendment to the Sherman Act made criminal convictions under Section 1 felony offenses and set the maximum prison sentence at 3 years. However, actual periods of incarceration for Sherman Act violations are usually less than 1 year. Convictions based on the Clayton Act are civil rather than criminal offenses and do not involve prison sentences.

Treble Damages Both the Sherman and the Clayton acts include provision for award of treble (triple) damages. If a private party can demonstrate that the antitrust laws have been broken and can prove the amount of damages sustained, the offending firm may be required to pay the plaintiff three times the amount of damages. Assume that Johnson, Inc. and Mack Manufacturing are convicted of price fixing by the government. Firms that purchased the products of these firms at inflated prices could sue for damages. A successful government suit is prima facie evidence that the firms had violated the law. Thus, the task of the plaintiffs would be to show the amount

of damages. Suppose that it is determined that a specific firm purchased $1,000,000 in supplies from Johnson, Inc. during the price-fixing period and that the total price would have been $900,000 in the absence of collusion. Thus, the plaintiff has been overcharged by $100,000 and would be entitled to claim three times that amount, or $300,000. The prospect of treble damage awards may be the most important deterrent to antitrust violations.

Key Concepts

- The courts have consistently ruled that price fixing is a per se violation of the antitrust laws.
- Predatory pricing is a means of weakening competitors and deterring entry into an industry. Recent court decisions have defined predation as existing if price is set below average variable cost.
- Most antitrust suits are settled by consent decrees.
- Cases that go to court may result in dissolution or divestiture, injunctions, fines, imprisonment, and/or the award of treble damages.

Case Study

Collusion in the Ivory Towers of Academia

In August 1989, the Department of Justice initiated an investigation involving 23 of the most prestigious private colleges in the United States. The allegation was that they were engaged in a conspiracy to raise tuition and to limit financial aid offered to students.

With respect to tuition, the charge was that the institutions shared information with one another regarding proposed increases. For example, an administrator at Harvard might have informed his counterpart at Yale that Harvard was contemplating raising tuition by 6 percent for the next year. Yale would use this information and similar data from the other Ivy League schools to determine its tuition rate. The result was that tuition rates were very similar. For 1989–1990, tuition, fees, and room and board at Yale were $19,310, while at Harvard they were just $85 more, at $19,395. Comparable totals at Dartmouth, Columbia, and the University of Pennsylvania were between those at Harvard and Yale.

The financial aid issue was somewhat more complicated. For many years, representatives of the nation's elite universities had met to coordinate the amount of assistance they would offer to outstanding students who had applied to more than one of the institutions. As a result of these discussions, students would be offered essentially identical amounts of aid from each school. The schools defended this practice as a way of allowing students to make their choices based on educational objectives rather than financial considerations. They also argued that it prevented bidding wars and allowed

limited financial aid dollars to be spread among a greater number of deserving students. Although these are laudable goals, another effect of the practice was that it reduced the amount of aid offered to top students.

In 1991, as part of a consent decree, the universities being investigated agreed to refrain from sharing tuition information, and the financial aid discussions were not held for the first time in 35 years. Although the schools avoided prosecution by the federal government, they were still vulnerable to private suits by individuals who were injured by their practices. ■

PUBLIC UTILITY REGULATION

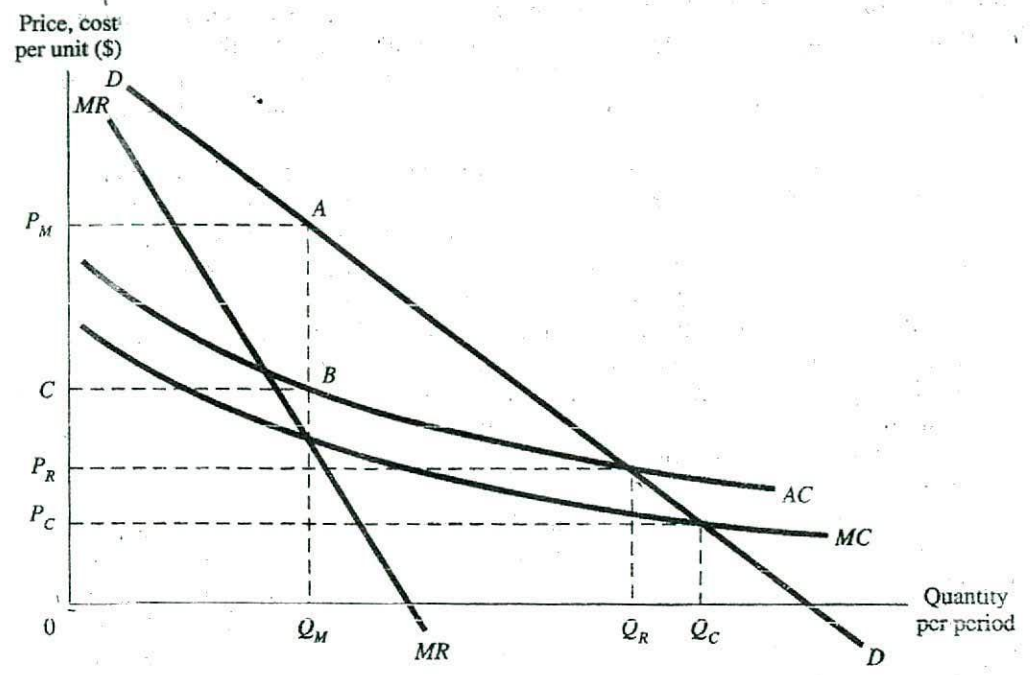
In contrast to antitrust activity, the regulatory approach to public policy directly alters industry performance by setting prices and establishing conditions of entry and exit. Although regulation extends to many industries, it is most common in dealing with public utilities. It is difficult to define precisely what makes a business a public utility. However, there seem to be two general characteristics. First, the industry provides a product or service of particular importance. Either the day-to-day livelihood or the future growth of a region depends on the continued and reasonable provision of the product or service. Second, the nature of the production process is such that competition is seen as yielding undesirable results such as duplication of facilities. The public utility designation is usually applied to firms providing electric power; local water and sewage supply; telephone, telegraph, and cable communications; and urban passenger transportation.

The Need for Regulation

The traditional view of public utility regulation is that it serves the public interest by protecting consumers. The need for regulation exists when a supplier has a natural monopoly or to prevent price discrimination.

Natural Monopoly Certain industries are sometimes referred to as *natural monopolies.* The term is used to describe production systems where technology results in continually declining average costs that provide a substantial cost advantage to larger firms. Because a firm can decrease its average cost by increasing its rate of output, the only stable market structure is that of a single firm serving the entire market. Smaller firms are either forced from the market or acquired by the dominant firm. Consumers in the market are then subject to the economic power of the resulting monopolist. Figure 19.3 depicts a natural monopoly. Note that the shape of the cost curves implies that a larger firm will be more efficient than a smaller one because it will have lower average costs at the optimal rate of output.

The existence of a natural monopoly poses something of a dilemma for public policy. One alternative is to let the firm operate as a monopoly. If the firm faced the demand curve DD, as shown in Figure 19.3, the monopoly price would be P_M and the quantity, Q_M. The firm would then earn economic profit, as indicated by the area of the



rectangle P_MABC. Compared to marginal cost pricing (i.e., setting the price equal to marginal cost), the monopoly-pricing scheme would result in a deadweight loss and also a transfer of consumer surplus from consumers to producers.[5]

If the firm is not allowed to act as an unconstrained monopolist, there are several alternatives available to the policymaker. One is to invoke antitrust laws and divide the firm into smaller competing firms. But it is not clear that the public interest would be served by this action. The inefficiency of these small firms (because of their higher average costs at lower output rates) requires that they charge a high price just to earn a normal return on capital. There is no guarantee that the price required by such firms to earn a normal return would not be higher than the price that the more efficient monopolist would charge to maximize profits. Thus, the antitrust approach may not be a desirable solution.

A second alternative is to allow the firm to maintain its monopoly position but require it to price at marginal cost. For the demand and cost curves in Figure 19.3, this would result in a price of P_C and a quantity of Q_C. At this level of production there is no deadweight loss because production is increased until the cost of producing the last unit is equated to the value of that unit. There is also no transfer of surplus from consumers to producers. In fact, the problem is quite the reverse. Be-

[5]Deadweight loss and consumer surplus are discussed on pages 328–329.

cause the monopolist is producing in a region of decreasing costs, its marginal cost is less than its average cost. Being required to price at marginal cost, the monopolist is unable to earn a normal return on capital. This is easily seen by observing that at the output rate Q_C, the average revenue as shown by the demand curve is less than the average cost.

If a policymaker requires a firm with decreasing costs to price at marginal cost over a long period of time, some provision must be made to compensate the investors in the firm for the losses that will be sustained. One way of doing this is to provide a subsidy. This approach is sometimes used with publicly owned bridges. The marginal cost of allowing another car over the bridge approaches zero. Thus, cars are permitted to pass without charge or at nominal cost, and the cost of building and operating the bridge is paid from tax revenues. In the United States, providing explicit subsidies to business from public funds has never been very popular. There have been exceptions, such as subsidies for airlines, urban transportation, and telephone service to rural customers, but the general philosophy has been that public utility services should not be subsidized.

The most common method for pricing the products of a natural monopoly in the United States represents a compromise solution. The nature of the compromise is depicted by Figure 19.3. A simple description of public utility price regulation is that price is set equal to average cost. That is, the firm is allowed to charge a price that allows it to earn no more than a normal return on its capital.[6] This is shown in Figure 19.3 by the price, P_R, and the quantity, Q_R. The regulatory approach is a compromise because the price is less than the monopoly price but higher than marginal cost. There is some deadweight loss because price is not equated to marginal cost, but the deadweight loss is far less than if the firm were allowed to act as a monopolist. Because the firm earns a normal profit, there is no need for the subsidy that would be required with marginal cost pricing. Thus, this mechanism achieves some of the gains from marginal cost pricing without requiring a subsidy.

Undue Price Discrimination Price discrimination occurs when consumers are charged different prices for a product and the differences in price cannot be accounted for by cost differentials. As discussed in chapter 12, the three requirements for successful price discrimination are that consumers have different demand elasticities, that markets be separable, and that the firm has some power over price.

The telephone industry provides an example of successful price discrimination policies by a public utility. Rates for basic telephone services are higher for business users than they are for residential users. There is no particular reason to assume that the cost of installing and maintaining a phone in an office is different from putting one in a kitchen. There are, however, possible differences in demand elasticity for business versus home phone customers. Consider the case of a stockbroker. The vast majority of orders for the purchase or sale of stock come to the broker by phone. There is no way the business could be conducted without a phone. In contrast, if there is a neighbor's phone that can be used in an emergency, it is quite possible to get along without a telephone

[6] Recall that average cost includes a normal return to capital. Thus, if price is set equal to average cost, the firm will be earning a normal return.

in one's home. In economic terms, the stockbroker is said to have more inelastic demand for telephone service than does the residential customer.

The other conditions for price discrimination are also met in the telephone industry. Because there is a physical connection between the customer and the phone company, there is no way that low-cost home telephone service can be resold to a business customer. Also, if the stockbroker does not interconnect with the local phone company, there is no practical way to have access to customers calling in orders.

The consequence of price discrimination provides an argument for regulation. Perhaps government should intervene to protect the commercial user from an unfair situation. The issue is not one of efficiency, but of fairness. The presumption is that the monopolist should not be allowed to use its power to unduly discriminate against some consumers. Although some discrimination may be acceptable, government intervention may be necessary when that discrimination becomes excessive. But there is no clear definition of the distinction between due and undue discrimination. In the end, undue price discrimination is whatever the regulatory commissions or the courts determine it to be.

Contestable Markets

The need for regulation and the resulting regulatory compromise just discussed is based on the assumption that natural monopolists can exercise their market power to the detriment of consumers. However, it is possible to accept the premise that a single supplier will serve the market most efficiently and still reject the conclusion that the firm will be able to set its prices substantially in excess of costs. One constraint on market power is provided by contestable markets.

Markets are considered contestable if competitors can easily and quickly enter when prices exceed costs and can exit if operations become unprofitable. This theory focuses on sunk costs rather than on economies of scale as the source of market power. That is, if entry into a market involves substantial expenditures that cannot be recouped on exit, potential competitors will be less likely to risk entry than if few sunk costs are involved. For example, the primary reason that competition is not viable in the electric power industry is because of the extremely high cost of the local distribution network. Once this local network has been constructed, the existing firm has substantial sunk costs, but its variable costs of distributing power are low. As a result, entry by a competitor is not likely to take place. In contrast, entry is more likely to occur in the airline industry, because the sunk costs are less important. Although an airplane can cost tens of millions of dollars, an entrant may lease aircraft or purchase secondhand planes. If the firm is unsuccessful, there is a ready market for the aircraft. Also, the potentially most significant source of sunk costs—airport runways and gate facilities—are financed and constructed by the government. Because these facilities are leased by the airlines, few sunk costs are involved.

An implication of the theory of contestable markets is that public policy should focus on methods of reducing sunk costs. One option would be for the government to construct facilities and lease them to private firms, as is done with airlines. Another would be to require the owners of distribution facilities to open those facilities to all competitors on an equal basis.

Key Concepts

- Public utilities are firms that provide an essential service and that have limited competition.
- When there is a natural monopoly, the most efficient market structure may be a single firm serving the entire market.
- Public utility regulation as practiced in the United States is based on setting price equal to average cost.
- Regulation may be needed to protect consumers from high prices and undue price discrimination by firms with market power.
- The need for regulation is reduced if markets are contestable. Contestable markets are those that permit easy entry and exit in response to price changes.

Regulatory Procedures

Public utility regulation occurs at both state and federal levels. Federal regulation focuses primarily on wholesale and interstate transactions, while state regulatory activities concentrate on the intrastate retail market. Commissions responsible for public utility regulation typically consist of two basic divisions: the commission staff and the commissioners.

The commission staff is an adversary of the regulated firm in hearings before the commissioners. The staff must evaluate the evidence presented by the firm and make its own recommendations. Although consumer groups are becoming increasingly important in regulatory proceedings, it is still the commission staff that has the major responsibility for presenting the public's case to the commissioners. In contrast, the commissioners consider evidence and recommendations of the staff, the regulated firms, and other groups, and then formulate the commission's policies.

Probably the most visible function of the state and federal regulatory commissions is setting prices for the products and services of the industries they regulate. The procedure used for price determination is a quasijudicial exercise called the *rate case*. There are two basic objectives in the rate case. The first is to find a general level of rates or prices that will allow the firm to earn no more than a fair or normal return on its capital. This occurs in the revenue requirement phase of the proceeding. The second objective is considered during the rate structure phase and involves setting rates that do not unduly discriminate against any class of consumers.

The Revenue Requirement Phase of the Rate Case The rate case is concerned primarily with equity or fairness. If prices are raised and profits increase, stockholders of the firm benefit at the expense of consumers. If the firm is not allowed to raise prices or if the increase does not cover increased costs, the benefits to consumers from lower rates are obtained at the expense of lower returns to investors. If a general increase in profits to the firm is granted while keeping prices to some groups low, higher prices must be paid by other groups. The rate case is basically a zero-sum game in which one group can benefit only at the expense of others. Thus, it becomes an adversary proceeding, with each of the parties involved trying to get a bigger share of the pie.

The procedure for determining the general level of prices in a rate case is easily described in theory, if not in practice. The firm is to be granted overall revenues sufficient to allow it to earn just a fair return on its capital. The procedure can be reduced to the following simple equation:

$$RR = E + s(RB - DEP) \tag{19-1}$$

where

RR = revenue requirement
E = expenses
s = fair return on capital
RB = capital or the rate base
DEP = depreciation

Equation (19-1) specifies that total revenue allowed by the regulatory commission should be sufficient to allow the firm to cover its expenses plus earn a fair return on the depreciated value of its capital base. Essentially, the public utility rate case is a cost-plus form of price setting. Revenues are set to cover the firm's costs of operating plus an add-on as a return to capital. In the rate case, each of the components of equation (19-1) is determined. The commission must decide on the firm's allowable expenses, a fair rate of return on capital, the depreciated capital or rate base to which the fair rate is to be applied, and finally, the revenue requirement necessary to cover the sum of expenses plus return to capital.

The Rate Structure Phase of the Rate Case The emphasis of the public utility rate case has shifted over time. In the early days of regulation, the major area of controversy was the valuation of the rate base. Later, attention shifted to the fair rate of return. Until recent years, commissions paid relatively little attention to the structure of rates. The need for increases in total revenue often was met simply by adjusting all rates on a nearly proportionate basis. Commissions considered the structure of rates primarily in response to complaints received from specific groups of consumers. Basically, it was the firm that took the lead in determining what the structure of rates should be.

In recent years, much more attention has been paid to the structure of rates. As commissions have become more involved in determining the rate structure, they have required that firms provide additional information about the cost of serving individual categories of consumers. Often, the commission will request that the firm compute the rate of return being earned for each category of service or for each product that is provided. These data are then used by the commission in making rate adjustments. This is called *cost-of-service pricing*. For example, if a particular service or product is shown to be earning a very low rate of return under existing rates, the commission may approve a larger rate increase than for a service or product that is earning a higher rate of return. The commission's attempt to equalize rates of return is consistent with its mandate to prevent undue discrimination.

Key Concepts

- In a rate case, the regulatory commission must determine prices that generate revenues sufficient to cover expenses and allow the firm to earn a fair return on its rate base.
- In determining the structure of rates, most regulatory commissions use cost-of-service pricing, which focuses on equating the rate of return earned by providing service to each customer class.

Interest Groups and Regulation

Some scholars do not accept the idea that regulation effectively protects consumers. They believe that the effects of regulation can best be explained in terms of regulators responding to pressures from various groups who have an interest in the outcome of a decision. Stigler is a leading proponent of this view.[7] He considers the regulatory process as a means of redistributing wealth. This redistribution occurs as commissions make decisions that tend to favor one group over another. The power of regulatory commissions could be used to confer benefits on any segment of society, but Stigler argues that industry is more effective than consumers in affecting the outcomes of decisions. Thus, regulation usually will favor industry at the expense of the consumer. But the main contribution by Stigler is not his conclusion that regulatory decisions are biased toward industry. It is his analysis that concludes that some groups are always likely to be better represented in regulatory proceedings than others. His work has evolved into a general theory of how interest groups affect the outcomes of political decisions.[8]

Modern interest group theory starts from the premise that regulation can have important effects on the distribution of wealth. As a result, those who will be affected by regulatory decisions have an incentive to try and influence those decisions. This can be accomplished by means such as lobbying, providing data, hiring expert witnesses, and conducting publicity campaigns. The responsibility of regulators is to sift through the information received and differentiate between arguments that reflect only special interests and those that support the public interest as the policymakers perceive it.

Limited budgets and staff force regulators to make decisions at least partially on the basis of information that is supplied by various interest groups. Thus, those groups that do the best job of representing their interests are most likely to receive favorable treatment by the regulatory body. But such representation can be a time-consuming and expensive process. As a result, certain types of special-interest groups may have an advantage in regulatory proceedings. Factors affecting the degree to which a group will be well-represented in a regulatory decision include degree of self-interest, size, homogeneity, and uncertainty.

---

[7]G. J. Stigler, "The Theory of Economic Regulation," *Bell Journal of Economics and Management Science* (Spring 1971): 3–21.

[8]See R. G. Noll and B. M. Owen, eds., *The Political Economy of Regulation: Interest Groups in the Regulatory Process* (Washington, D.C.: American Enterprise Institute, 1983).

Degree of Self-Interest The more important a decision is to a group, the more effort the members of that group will expend to influence the decision. For example, an electric utility that is requesting a $100 million rate increase has a much greater stake in the decision of the regulatory commission than does the consumer whose monthly electricity bill will increase by $3.50 if the rate increase is approved. As a result, the utility will spend a great deal of time and effort trying to sway the commission, while an individual consumer may do little more than grumble when the higher bill comes.

Size of the Group All other thing being equal, a smaller group is better able to promote its views than a group consisting of many people or firms. The reason is that organizational problems increase with size. Also, large groups tend to have more serious problems with free riders. A single consumer, asked to contribute to a fund to lobby for reduced electricity rates, may rationalize that his or her failure to contribute will have no noticeable effect on the effectiveness of the group. In contrast, in a consortium of three or four firms, one defector might spell the difference between success and defeat.

Group Homogeneity If all of the members of a group are in a similar position and have similar objectives, then the group does not have to spend time and money hammering out compromise positions for presentation before the regulatory body. All of the resources of the group can be marshaled to achieve the common goal.

Uncertainty If the results of a particular decision are predictable, then members of a group will be more willing to make contributions than if the results are uncertain. For example, if there is a lack of information as to whether deregulation of an industry will lead to higher or lower prices for consumers, then consumer groups may have little impact on the decision process.

Stigler's early formulation of interest-group theory predicted that regulation is biased toward producers. The modern theory is roughly consistent, but somewhat more general. Existing firms in an industry are likely to have a high stake in regulatory decisions, be relatively few in number, hold similar objectives, and have a good idea about the outcome of regulatory decisions. Thus, the views of those firms are likely to be well-represented in the regulatory process. In contrast, consumers, small firms attempting to enter a regulated industry, and those advocating the use of new technologies are likely to be poorly represented. As a result, the theory predicts that there will be a tendency for regulatory decisions to favor the existing firms in a regulated industry. The theory also suggests that small, one-issue consumer groups such as environmentalists and senior citizens can have an important impact on the outcomes of decisions.

Key Concepts

- The regulatory process can be used as a means of redistributing wealth.
- Interest-group theory predicts that small, homogeneous groups that are significantly affected by a regulatory decision are likely to be favored by those decisions.

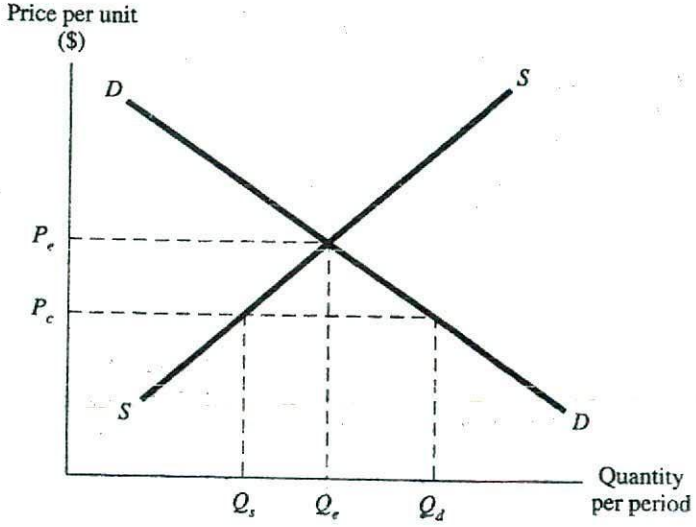
PRICE CONTROLS

During periods of rapid inflation, price controls may be used to reduce the rate of price increases. In some cases, they are imposed on selected industries (such as rent controls or interest rate ceilings), while in other circumstances they have been used in virtually all industries.

Price Controls in Competitive Industries

Figure 19.4 depicts the supply and demand curves for a competitive industry. Each firm is considered to take as given the market price determined by the interaction of the supply and demand curves. The equilibrium price is P_e and the equilibrium quantity is Q_e. Suppose that public policy limits the price at which the product can be sold to P_c. The lower price reduces the amount that will be supplied to Q_s and induces additional consumers to enter the market so that the quantity demanded increases to Q_d. The amount that consumers want to purchase is now greater than the available supply, so there is a shortage of the price-controlled product.

In the free market, prices act as a rationing device to equilibrate supply and demand. Some consumers drop out of the market as prices increase. At the same time, quantity supplied will increase as production becomes more profitable. But when prices are set below the equilibrium level, another method of rationing must be found. Historically, alternative rationing schemes have taken many forms. The most common has been the black market. If products can be purchased at low prices by consumers who value them less than others who cannot obtain them, an opportunity to profit from exchange exists. The low-valuation consumers can sell the product to the high-valuation consumers, and both will be better off than if the black-market transaction had not occurred. The existence of black markets requires that there be people who are willing to break the law and risk incurring the penalties of running a black-market supply orga-



nization. It also requires that there be some consumers who are willing to purchase goods illegally. There has never been a documented case of effective price ceilings that was not accompanied by some sort of black market.

In many countries, the common manifestation of prices artificially kept below the market level is long queues of people waiting to purchase the limited supplies of goods. Such queues are common in poor nations and have periodically occurred at gas stations in the United States when price controls were in effect. Queues are a form of rationing in the sense that only those people willing and able to stand in line get the product. The inefficiency of this type of rationing is apparent when the value of time is considered. Suppose that the price of a product as determined by market forces would be $2, but the controlled price is set at $1. If the purchaser's time is valued at, say, $5 an hour, he has to stand in line only 12 minutes per unit purchased to lose the gain of being able to buy at a lower price.

Thus far, only the short-run implications of price controls of firms in competitive markets have been considered. The long-run impacts on investment may be far more important. One of the functions of prices is to signal needs for transferring resources from one sector of the economy to another. If prices are rising because of demand-induced pressures in one industry, capital tends to flow to that industry to increase supply. Rising prices also generate internal investment funds for expansion by firms within the industry. But if prices are kept at artificially low levels, the necessary signals are not provided to capital markets, and sufficient internal funds are not generated. As a result, if controls are imposed for a long period of time, shortages become more acute because of the lack of expansion in the industry. When controls are lifted, pressures of excess demand may result in a significant price increase.

Case Study

Rent Controls in Paris

Rent controls have been imposed by many city governments. In almost every case, the result has been a shortage of rental units and a deterioration in the quality of rental housing. A dramatic example is Paris after World War II. Those lucky enough to have rent-controlled housing in Paris seldom had to pay more than 4 percent of their income for housing. Today, it is not uncommon for housing costs to take 30 percent of a family's gross income.

Unfortunately, quantity demanded far exceeded quantity supplied. For those who did not already have a low-rent unit, there were none available. The death of someone living in a rent-controlled unit was about the only possibility. Search for an apartment sometimes involved reading the obituaries or making agreements with undertakers for early notification of a death. Old people were often accosted by young wives wanting to make a "down payment" on future space. The rights to a rent-controlled apartment were sold for as much as $6,000 (in terms of 1998 dollars) per room. This high price is an indication of the misallocation of resources caused by the rent controls.

Another consequence of rent controls in Paris was the deterioration of the rental housing stock. With rents so low, it did not pay to build additional units, nor did it make sense to spend very much to maintain existing units. Between 1914 and 1948, the increase in rental rates was 6.8 times, while the cost of repairs increased by 120 times. As a result, repairs and other maintenance were neglected. ■

Price Controls and Firms with Market Power

On the basis of the theoretical discussions of the previous pages, the attempt to apply controls to competitive industries is not very promising. However, even casual observation of business in any developed nation reveals that many goods and services are not produced under conditions that satisfy or even approximate the perfectly competitive model. Rather, much of the economic activity in industrialized nations involve firms with some power to set the price of their product. This power is not complete because such firms have rivals, but it sometimes is enhanced by formal or informal price collusion among firms.

Where firms have power to affect prices, some of the traditional objections to price controls become less compelling. For economists who express concern about the loss of freedom resulting from price controls, a relevant question in this case is "Whose freedom?" For firms with market power, the invisible hand no longer functions as envisioned by Adam Smith. The loss of freedom caused by price controls is that of firms who have been exploiting the consumers of their product. For such firms, price controls might be viewed as limits on operations that markets do not effectively constrain.

Figure 19.5 depicts the application of price controls to a firm with market power. If left to maximize profits, the firm sets price at P_m and produces the quantity Q_m. By setting a price ceiling where the marginal cost curve crosses the demand schedule (point A), it is possible to constrain price without creating a shortage. The profit-maximizing firm chooses a quantity such that marginal revenues and marginal costs are equal. If the maximum price allowed is P_c, the firm has a new marginal revenue curve, P_cA, out to quantity Q_c. Because P_c is below the demand curve for all quantities less than Q_c, the firm can increase production from zero to Q_c by selling additional units at the same price. That is, the marginal revenue curve is given by the horizontal straight line P_cA. The firm should continue to increase its production as long as marginal costs are less than P_c. Thus, the profit-maximizing quantity is Q_c. If price is set at P_c, there is no shortage or excess. The firm maximizes profits by producing Q_c, which is the same amount as consumers want to purchase at the controlled price.

It should be noted, however, that analyzing price controls is much easier than their actual implementation. Regulators seldom have good information on demand and costs. Thus, it would be unlikely that policymakers could consistently identify the precise level at which the demand and marginal cost curves intersect. If price is set below this level, shortages will result. A price above that level means that prices are higher than necessary to equate demand and supply.

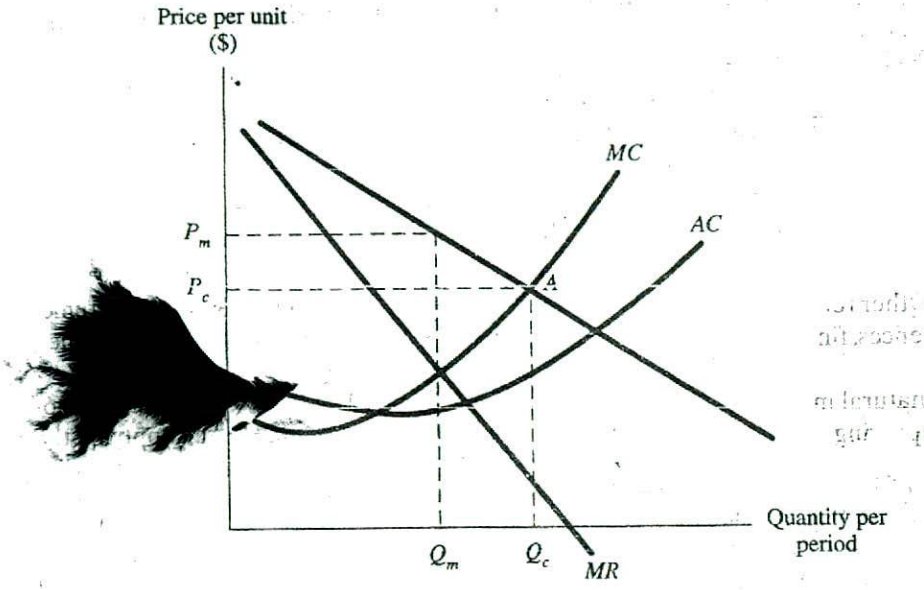


FIGURE 19-5 Price Controls in Concentrated Industries

Key Concepts

- Price controls in competitive industries tend to result in shortages.
- In industries where firms have market power, if price is set equal to marginal cost, it may be possible to impose price controls without creating shortages.

SUMMARY

The Sherman Act prohibits unfair business practices and the illegal monopolization of a market. Price discrimination and mergers are two important actions covered by the Clayton Act. The Robinson–Patmen Act broadened Clayton Act provisions against illegal price discrimination, while the Celler–Kefauver Amendment put vertical and conglomerate mergers under the jurisdiction of the Clayton Act.

Antitrust suits may be initiated by the U.S. Department of Justice, the Federal Trade Commission, state government officials, or private parties. For successful prosecution, per se offenses require only showing that the act has been committed. In contrast, rule-of-reason offenses also require a demonstration that the social costs of the act are greater than the social benefits.

The current interpretation of Section 2 of the Sherman Act is that practices that are acceptable when used by small firms may constitute antitrust violations if used by firms with substantial market power. Mergers contested under the Clayton Act should be evaluated to determine if the effect of increased market power exceeds any

efficiency gains that may result from the merger. Income distribution effects may also be a consideration.

In deciding to challenge a horizontal merger, the Justice Department may rely on guidelines based on the Herfindahl Index. A horizontal merger between large firms is likely to be prohibited, but vertical and conglomerate mergers will probably be allowed unless they increase horizontal market power.

Price fixing is usually considered a per se violation of the law. Predatory pricing can be used to weaken competitors and deter entry. Recent court opinions have defined predation as setting prices below average variable cost.

Most antitrust suits are settled by consent decrees before going to trial. [illegible]eme-dies and penalties include dissolution or divestiture, injunctions, prison s[illegible]es, and treble-damage awards.

Firms with continually declining average cost curves are refe[illegible]o-nopolies because the only stable market structure is that of a single firm [illegible]pply[illegible] the entire market. In the United States, the approach used to deal with natural monopolies is to give price-setting power to a regulatory body. Such bodies also protect consumers by prohibiting undue price discrimination. The need for regulation is reduced if markets are contestable. Contestable markets are those that permit easy entry and exit in response to price changes.

In the revenue requirement phase of a rate case, a regulatory commission must determine the total revenue that will allow the firm to cover its expenses plus earn a fair return on its rate base. In the rate structure phase, a cost-of-service approach is often used whereby the commission adjusts rates to equalize the rate of return earned by each class of service. Modern interest group theory predicts that decisions of regulatory bodies will tend to favor small, homogeneous groups that will be significantly affected by the decision.

Price controls in competitive industries tend to cause shortages and result in queues and black markets. In contrast, in industries where firms have market power, by setting the controlled price equal to marginal cost, it may be possible to impose price controls without creating shortages.

cussion Questions

19-1. How could private antitrust suits be used as part of a firm's strategy in dealing with its competitors?

19-2. Why are some antitrust violations judged using a per se standard?

19-3. How could the definition of the relevant product market affect a court's decision in an antitrust case?

19-4. Consider an industry with a large, dominant firm and a number of smaller firms. Suppose that a merger of two of the small firms would result in a Herfindahl Index greater than 1,800. Is there a basis for allowing the merger?

19-5. Trade associations are often involved in collusive agreements to fix prices. Why?

19-6. In the Trenton Potteries case, what did the Supreme Court mean by the statement "the reasonable price fixed today may through economic and business changes become the unreasonable price of tomorrow"?

19-7. How would the value of a good relative to its transportation cost affect the success of a large, multimarket firm using predatory pricing? Explain.

19-8. Can a monopoly be maintained in an industry that does not have continually declining average costs? Explain.

19-9. In a public utility rate case, how could a commission use the structure of rates to redistribute income? Give an example.

19-10. How might price regulation affect the rate of innovation in an industry?

19-11. How should a regulatory commission determine what constitutes undue price discrimination?

19-12. In competitive industries, when prices are kept below the equilibrium level by price controls, incorrect signals are provided to consumers and resource owners. Explain.

19-13. How do elasticities of supply and demand affect the magnitude of the shortages created by price controls imposed on competitive industries?

19-14. Would controls be more successful in keeping prices down during wartime than during a period of peace? Why or why not?

Problems

19-1. Bentley Manufacturing can produce desks at a constant average cost of $200 per unit, including a normal profit. Because of intense competition, the company sells desks for $200 and is able to sell 100,000 units per year. Bentley's managers have initiated a merger with New Top, another desk manufacturer. The merger will allow the combined firm to produce at an average cost of $190 per unit. The combined firm's additional market power increases the profit-maximizing price to $250, but only 80,000 units will be sold at that price. The proposed merger is being evaluated by the Department of Justice.

a. If the demand curve is linear between a price of $150 and $250, will the deadweight loss exceed the efficiency gain resulting from the merger?

b. What other factors should antitrust officials consider in deciding whether to challenge the merger?

19-2. An industry consists of eight firms. Sales for each firm in 1988 and 1998 are shown. Use the Herfindahl Index to determine whether the industry became more concentrated from 1988 to 1998.

	Sales (millions)	
Firm	***1988***	***1998***
1	100	240
2	90	90
3	80	60
4	80	50
5	50	50
6	40	40
7	30	40
8	30	30
Total	500	600

19-3. Market shares for the firms in two industries are as follows:

Firm	*Industry I*	*Industry II*
1	5%	5%
2	5	5
3	20	5
4	5	60
5	5	2
6	20	3
7	20	5
8	20	15

Compute the Herfindahl Index for each industry.

19-4. An industry consists of eight firms of equal size. Two of the eight want to merge. Based on the merger guidelines used by the Department of Justice, will the merger be challenged? What if the merger involved four of the eight firms in the industry? Explain your answers.

19-5. A firm's demand function is given by $P = 100 - 4Q$. Marginal costs are constant and equal to \$20. Competitive conditions in the industry require the firm to price at marginal cost. Scale economies resulting from merging with a competing firm would reduce the marginal cost to \$19 and allow the firm to increase its price to \$28. Use the Williamson model to evaluate the merger.

19-6. Malko Electronics is a manufacturer and distributor of Korean-made video recorders. The firm's most popular model sells for \$250 in California. Shipping costs from Korea to the United States are \$10 per unit. Fixed costs plus a normal rate of profit are \$100 per recorder, and the average variable cost is \$170. A competitor, Wagner Video, sells a comparable recorder in California for \$280. Wagner has filed a price discrimination suit, charging that Malko is engaged in predatory pricing intended to drive Wagner from the market. The charge is based solely on the contention that Malko is selling below cost. What will be the important issues in the case? Is Malko likely to be found guilty of illegal price discrimination?

19-7. A regulated telephone company utility has a depreciated rate base of \$500 million. The utility's capital structure consists of 40 percent equity and 60 percent debt. The after-tax cost of debt is 8 percent, the firm's common stock is currently selling for \$8 per share, the dividend is \$1, and the expected dividend growth rate is 2.5 percent per year. The firm's total expenses are \$400 million.

a. If the regulatory body sets an allowed rate of return that is equal to the firm's cost of capital, what rate of return will the commission allow? Assume the commission uses the discounted cash flow method to compute the cost of equity capital.

b. Based on your answer to part (a), what is the revenue requirement of the firm?

19-8. Pilfur Oil has a monopoly on the sale of oil in Transylvania. Demand for the firm's product is given by

$$P = 100 - 20Q$$

where P is the price of a barrel of oil and Q is measured in thousands of barrels of oil per day. The marginal cost function for the firm is

$$MC = 10 + 5Q$$

a. What are the profit-maximizing price and rate of output for Pilfur?
b. The government of Transylvania decides that Pilfur's price is too high. The firm is required to set the lowest price that will not result in shortages. What will be the price and rate of output that meet these conditions?

19-9. The supply and demand equations for a competitive industry are given by:

$$\text{Supply:} \quad Q = 100 - 4P$$
$$\text{Demand:} \quad Q = 50 + 5P$$

a. If the government mandates that the maximum price is \$4, what will be the price and the quantity exchanged in the market?
b. If the maximum price is \$6, what will be the price and the quantity exchanged in the market?

Integrating CASE Study VIII

Autovideo (Un), Ltd.

Three years ago Autovideo, Ltd. developed a new method of renting videocassettes. Basically, the idea involves distribution through automated outlets in shopping centers. To use the service, customers must first obtain a membership that includes a customer identification number. There is no cost for the membership, but it is issued only after the company receives verification that the individual is of legal age. Once received, the ID number is used in the following way to obtain a cassette. The customer enters the ID number at a terminal in a rental outlet. If a valid number is entered, a list of available cassettes is displayed. To obtain a cassette, the customer simply enters the number of the cassette. This action causes the cassette to be given to the customer and a record to be made of the transaction. When a cassette is returned, the rental amount is recorded on the person's account. Each month the customer is sent a bill for rentals during that period.

Autovideo was started in South Dakota, and the company has a dominant share of the market for cassette rentals in that state. The firm's success is the result of being first to offer the service. Although there are nonautomated establishments that provide movies, they presently offer neither the convenience nor the range of choice of Autovideo. In addition, the firm greatly benefitted from publicity caused by vocal opposition from local politicians, who were concerned about the effect that a large firm like Autovideo might have on the smaller businesses.

Rental prices are $5 per movie in South Dakota. During its last fiscal year, the firm's South Dakota operations earned total revenue of over $10,000,000 and a pretax profit of 50 percent.

MovieView began using Autovideo's method 15 months ago in North Dakota. Autovideo immediately proposed a merger, but the offer was rejected by the management of MovieView. A year ago, Autovideo began opening rental outlets of their own in North Dakota. First year sales totaled $500,000, and the two firms currently each have half of the total market. In an effort to compete more vigorously, Autovideo recently decreased its price in North Dakota to $2.50 per movie. After the price change, Autovideo's North Dakota outlets averaged 100 rentals per day.

To facilitate better decision making, an accountant was hired by Autovideo to collect and analyze cost information. He determines that scale economies are unimportant and that daily costs for a typical outlet are as follows:

LEASE PAYMENTS: $40 per day. In each case, the lease contract is for 2 years and the property cannot be sublet to anyone else.

MOVIE INVENTORY: $40 per day. Provisions of the purchase contract prevent resale to other firms or to the general public. Because of duplication of cassettes, an outlet's inventory cannot be used efficiently at other outlets.

COMPUTER COSTS: $100 per day. The system is operated by use of a time-shared computer. The agreement with the system owner is based on actual days used and may be canceled at any time without penalty.

RENTAL COSTS: $1.00 per rental. The only cost that varies with the number of rentals is the billing cost and the replacement cost of cassettes. The more often a cassette is rented, the sooner it has to be replaced.

At a trade meeting, it is reported that the demand for movie rentals is very inelastic. After the meeting, Autovideo managers return to their office to find that:

1. In response to a complaint from MovieView, the North Dakota Attorney General's office is investigating alleged predatory pricing of cassettes by Autovideo in the state.
2. A South Dakota state legislator appeared on a talk show to condemn the market of Autovideo. Noting the firm's 50 percent pretax profit in the state, the legislator proposed a flat-rate tax on profits earned from rental of cassettes. He believes the tax would increase competition and reduce Autovideo's market share.
3. At a news conference, the chairman of the tax commission in South Dakota proposes a $0.50 excise tax on the rental of cassettes. He argues that such tax would be a good way to raise revenue.
4. In a Labor Day speech at Mt. Rushmore, a past governor of South Dakota decries the market power of large firms in the videocassette rental industry and announces that she supports the creation of a nonprofit enterprise to be given a monopoly right to sell and rent these materials. This enterprise would be government-owned and similar to state-owned liquor stores that exist in many states. The main purpose of the enterprise would be to keep prices down.

Requirements

1. As a law and economics professor in North Dakota, you are interested in the predatory pricing charge. Prepare a report that analyzes the strengths and weaknesses of this allegation.
2. The present governor of South Dakota must respond to the proposals for a profit tax, excise tax, and government ownership. You are the governor's administrative assistant. Prepare a report that analyzes the probable effects of each proposal.

Tables

TABLE 1 Present Value of 1

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

Periods (n)	Interest (discount) Rate (i)									
	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
1	.9901	.9804	.9709	.9615	.9524	.9434	.9346	.9259	.9174	.9091
2	.9803	.9612	.9426	.9246	.9070	.8900	.8734	.8573	.8417	.8264
3	.9706	.9423	.9151	.8890	.8638	.8396	.8163	.7938	.7722	.7513
4	.9610	.9238	.8885	.8548	.8227	.7921	.7629	.7350	.7084	.6830
5	.9215	.9057	.8626	.8219	.7835	.7473	.7130	.6806	.6499	.6209
6	.9420	.8880	.8375	.7903	.7462	.7050	.6663	.6302	.5963	.5645
7	.9327	.8706	.8131	.7599	.7107	.6651	.6227	.5835	.5470	.5132
8	.9235	.8536	.7894	.7307	.6768	.6274	.5820	.5403	.5019	.4665
9	.9143	.8368	.7664	.7026	.6446	.5919	.5439	.5002	.4604	.4241
10	.9053	.8203	.7441	.6756	.6139	.5584	.5083	.4632	.4224	.3855
11	.8963	.8043	.7224	.6496	.5847	.5268	.4751	.4289	.3875	.3505
12	.8874	.7885	.7014	.6246	.5568	.4970	.4440	.3971	.3555	.3186
13	.8787	.7730	.6810	.6606	.5303	.4688	.4150	.3677	.3262	.2897
14	.8700	.7579	.6611	.5775	.5051	.4423	.3878	.3405	.2992	.2533
15	.8613	.7430	.6419	.5553	.4810	.4173	.3624	.3152	.2745	.2394
16	.8528	.7284	.6232	.5339	.4581	.3936	.3387	.2919	.2519	.2176
17	.8444	.7142	.6050	.5134	.4363	.3714	.3166	.2703	.2311	.1978
18	.8360	.7002	.5874	.4936	.4155	.3503	.2959	.2502	.2120	.1799
19	.8277	.6864	.5703	.4746	.3957	.3305	.2765	.2317	.1945	.1635
20	.8195	.6730	.5537	.4564	.3769	.3118	.2584	.2145	.1784	.1486
21	.8114	.6598	.5375	.4388	.3589	.2942	.2415	.1987	.1637	.1351
22	.8034	.6468	.5219	.4220	.3418	.2775	.2257	.1839	.1502	.1228
23	.7954	.6342	.5067	.4057	.3256	.2618	.2109	.1703	.1378	.1117
24	.7876	.6217	.4919	.3901	.3101	.2470	.1971	.1577	.1264	.1015
25	.7798	.6095	.4776	.3751	.2953	.2330	.1842	.1460	.1160	.0923
26	.7720	.5976	.4637	.3607	.2812	.2198	.1722	.1352	.1064	.0839
27	.7644	.5859	.4502	.3468	.2678	.2074	.1609	.1252	.0976	.0763
28	.7568	.5744	.4371	.3335	.2551	.1956	.1504	.1159	.0895	.0693
29	.7493	.5631	.4243	.3207	.2429	.1846	.1406	.1073	.0822	.0630
30	.7419	.5521	.4120	.3083	.2314	.1741	.1314	.0994	.0754	.0573
35	.7059	.5000	.3554	.2534	.1813	.1301	.0937	.0676	.0490	.0356
40	.6717	.4529	.3066	.2083	.1420	.0972	.0668	.0460	.0318	.0221
45	.6391	.4102	.2644	.1712	.1113	.0727	.0476	.0313	.0207	.0137
50	.6080	.3715	.2281	.1407	.0872	.0543	.0339	.0213	.0134	.0085
55	.5785	.3365	.1968	.1157	.0683	.0406	.0242	.0145	.0087	.0053

TABLE 1 (continued)

Periods (n)	Interest (discount) Rate (i)									
	12%	14%	15%	16%	18%	20%	24%	28%	32%	36%
1	.8929	.8772	.8696	.8621	.8475	.8333	.8065	.7813	.7576	.7353
2	.7972	.7695	.7561	.7432	.7182	.6944	.6504	.6104	.5739	.5407
3	.7118	.6750	.6575	.6407	.6086	.5787	.5245	.4768	.4348	.3975
4	.6355	.5921	.5718	.5523	.5158	.4823	.4230	.3725	.3294	.2923
5	.5674	.5194	.4972	.4761	.4371	.4019	.3411	.2910	.2495	.2149
6	.5066	.4556	.4323	.4104	.3704	.3349	.2751	.2274	.1890	.1580
7	.4523	.3996	.3759	.3538	.3139	.2791	.2218	.1776	.1432	.1162
8	.4039	.3506	.3269	.3050	.2660	.2326	.1789	.1388	.1085	.0854
9	.3606	.3075	.2843	.2630	.2255	.1938	.1443	.1084	.0822	.0628
10	.3220	.2697	.2472	.2267	.1911	.1615	.1164	.0847	.0623	.0462
11	.2875	.2366	.2149	.1954	.1619	.1346	.0938	.0662	.0472	.0340
12	.2567	.2076	.1869	.1685	.1372	.1122	.0757	.0517	.0357	.0250
13	.2292	.1821	.1625	.1452	.1163	.0935	.0610	.0404	.0271	.0184
14	.2046	.1597	.1413	.1252	.0985	.0779	.0492	.0316	.0205	.0135
15	.1827	.1401	.1229	.1079	.0835	.0649	.0397	.0247	.0155	.0099
16	.1631	.1229	.1069	.0930	.0708	.0541	.0320	.0193	.0118	.0073
17	.1456	.1078	.0929	.0802	.0600	.0451	.0258	.0150	.0089	.0054
18	.1300	.0946	.0808	.0691	.0508	.0376	.0208	.0118	.0068	.0039
19	.1161	.0829	.0703	.0596	.0431	.0313	.0168	.0092	.0051	.0029
20	.1037	.0728	.0611	.0514	.0365	.0261	.0135	.0072	.0039	.0021
21	.0926	.0638	.0531	.0443	.0309	.0217	.0109	.0056	.0029	.0016
22	.0826	.0560	.0462	.0382	.0262	.0181	.0088	.0044	.0022	.0012
23	.0738	.0491	.0402	.0329	.0222	.0151	.0071	.0034	.0017	.0008
24	.0659	.0431	.0349	.0284	.0188	.0126	.0057	.0027	.0013	.0006
25	.0588	.0378	.0304	.0245	.0160	.0105	.0046	.0021	.0010	.0005
26	.0525	.0331	.0264	.0211	.0135	.0087	.0037	.0016	.0007	.0003
27	.0469	.0291	.0230	.0182	.0115	.0073	.0030	.0013	.0006	.0002
28	.0419	.0255	.0200	.0157	.0097	.0061	.0024	.0010	.0004	.0002
29	.0374	.0224	.0174	.0135	.0082	.0051	.0020	.0008	.0003	.0001
30	.0334	.0196	.0151	.0116	.0070	.0042	.0016	.0006	.0002	.0001
35	.0189	.0102	.0075	.0055	.0030	.0017	.0005	.0002	.0001	*
40	.0107	.0053	.0037	.0026	.0013	.0007	.0002	.0001	*	*
45	.0061	.0027	.0019	.0013	.0006	.0003	.0001	*	*	*
50	.0035	.0014	.0009	.0006	.0003	.0001	*	*	*	*
55	.0020	.0007	.0005	.0003	.0001	*	*	*	*	*

*$PVIF < .00005$.

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

Periods (n)	Interest (discount) Rate (i)								
	1%	2%	3%	4%	5%	6%	7%	8%	9%
1	0.9901	0.9804	0.9709	0.9615	0.9524	0.9434	0.9346	0.9259	0.9174
2	1.9704	1.9416	1.9135	1.8861	1.8594	1.8334	1.8080	1.7833	1.7591
3	2.9410	2.8839	2.8286	2.7751	2.7532	2.6730	2.6243	2.5771	2.5313
4	3.9020	3.8077	3.7171	3.6299	3.5460	3.4651	3.3872	3.3121	3.2397
5	4.8634	4.7135	4.5797	4.4518	4.3295	4.2124	4.1002	3.9927	3.8897
6	5.7955	5.6014	5.4172	5.2421	5.0757	4.9173	4.7665	4.6229	4.4859
7	6.7282	6.4720	6.2303	6.0021	5.7894	5.5824	5.3893	5.2064	5.0030
8	7.6517	7.3255	7.0197	6.7327	6.4632	6.2098	5.9713	5.7466	5.5348
9	8.5660	8.1622	7.7861	7.4353	7.1078	6.8017	6.5152	6.2469	5.9952
10	9.4713	8.9826	8.5302	8.1109	7.7217	7.3601	7.0236	6.7101	6.4177
11	10.3676	9.7868	9.2526	8.7605	8.3064	7.8869	7.4987	7.1390	6.8052
12	11.2551	10.5753	9.9540	9.3851	8.8633	8.3838	7.9427	7.5361	7.1607
13	12.1337	11.3484	10.6350	9.9856	9.3936	8.8527	8.3577	7.9038	7.4869
14	13.0037	12.1062	11.2961	10.5631	9.8986	9.2950	8.7455	8.2442	7.7862
15	13.8651	12.8493	11.9379	11.1184	10.3797	9.7122	9.1079	8.5595	8.0607
16	14.7179	13.5777	12.5611	11.6523	10.8378	10.1059	9.4466	8.8514	8.3126
17	15.5623	14.2919	13.1661	12.1657	11.2741	10.4773	9.7632	9.1216	8.5436
18	16.3983	14.9920	13.7535	12.6593	11.6896	10.8276	10.0591	9.3719	8.7556
19	17.2260	15.6785	14.3238	13.1339	12.0853	11.1581	10.3356	9.6036	8.9501
20	18.0456	16.3514	14.8775	13.5903	12.4622	11.4699	10.5940	9.8181	9.1285
21	18.8570	17.0112	15.4150	14.0292	12.8212	11.7641	10.8355	10.0186	9.2922
22	19.6604	17.6580	15.9369	14.4511	13.1630	12.0416	11.0612	10.2007	9.4424
23	20.4558	18.2922	16.4436	14.8568	13.4886	12.3034	11.2722	10.3711	9.5802
24	21.2434	18.9139	16.9355	15.2470	13.7986	12.5504	11.4693	10.5288	9.7066
25	20.0232	19.5234	17.4131	15.6221	14.0939	12.7834	11.6536	10.6748	9.8226
26	22.7952	20.1210	17.8768	15.9828	14.3752	13.0032	11.8258	10.8100	9.9290
27	23.5596	20.7069	18.3270	16.3296	14.6430	13.2105	11.9867	10.9352	10.0266
28	24.3164	21.2813	18.7641	16.6631	14.8981	13.4062	12.1371	11.0511	10.1161
29	25.0658	21.8444	19.1885	16.9837	15.1411	13.5907	12.2777	11.1584	10.1983
30	25.8077	22.3965	19.6004	17.2920	15.3725	13.7648	12.4090	11.2578	10.2737
35	29.4086	24.9986	21.4872	18.6646	16.3742	14.4982	12.9477	11.6546	10.5668
40	32.8347	27.3555	23.1148	19.7928	17.1591	15.0463	13.3317	11.9246	10.7574
45	36.0945	29.4902	24.5187	20.7200	17.7741	15.4558	13.6055	12.1084	10.8812
50	39.1961	31.4236	25.7298	21.4822	18.2559	15.7619	13.8007	12.2335	10.9617
55	42.1472	33.1748	26.7744	22.1086	18.6335	15.9905	13.9399	12.3186	11.0140

TABLE II *(continued)*

Periods (n)	Interest (discount) Rate (i)									
	10%	12%	14%	15%	16%	18%	20%	24%	28%	32%
1	0.9091	0.8929	0.8772	0.8696	0.8621	0.8475	0.8333	0.8065	0.7813	0.7576
2	1.7355	1.6901	1.6467	1.6257	1.6052	1.5656	1.5278	1.4568	1.3916	1.3315
3	2.4869	2.4018	2.3216	2.2832	2.2459	2.1743	2.1065	1.9813	1.8684	1.7663
4	3.1699	3.0373	2.9137	2.8550	2.7982	2.6901	2.5887	2.4043	2.2410	2.0957
5	3.7908	3.6048	3.4331	3.3522	3.2743	3.1272	2.9906	2.7454	2.5320	2.3452
6	4.3553	4.1114	3.8887	3.7845	3.6847	3.4976	3.3255	3.0205	2.7594	2.5342
7	4.8684	4.5638	4.2883	4.1604	4.0386	3.8115	3.6046	3.2423	2.9370	2.6775
8	5.3349	4.9676	4.6389	4.4873	4.3436	4.0776	3.8372	3.4212	3.0758	2.7860
9	5.7590	5.3282	4.9464	4.7716	4.6065	4.3030	4.0310	3.5655	3.1842	2.8681
10	6.1446	5.6502	5.2161	5.0188	4.8332	4.4941	4.1925	3.6918	3.2689	2.9304
11	6.4951	5.9377	5.4527	5.2337	5.0286	4.6560	4.3271	3.7757	3.3351	2.9776
12	6.8137	6.1944	5.6603	5.4206	5.1971	4.7932	4.4392	3.8514	3.3868	3.0133
13	7.1034	6.4235	5.8424	5.5831	5.3423	4.9095	4.5327	3.9124	3.4272	3.0404
14	7.3667	6.6282	6.0021	5.7245	5.4675	5.0081	4.6106	3.9616	3.4587	3.0609
15	7.6061	6.8109	6.1422	5.8474	5.5755	5.0916	4.6755	4.0013	3.4834	3.0764
16	7.8237	6.9740	6.2651	5.9542	5.6685	5.1624	4.7296	4.0333	3.5026	3.0882
17	8.0216	7.1196	6.3729	6.0472	5.7487	5.2223	4.7746	4.0591	3.5177	3.0971
18	8.2014	7.2497	6.4674	6.1280	5.8178	5.2732	4.8122	4.0799	3.5294	3.1039
19	8.3649	7.3658	6.5504	6.1982	5.8775	5.3162	4.8435	4.0967	3.5386	3.1090
20	8.5136	7.4694	6.6231	6.2593	5.9288	5.3527	4.8696	4.1103	3.5458	3.1129
21	8.6487	7.5620	6.6870	6.3125	5.9731	5.3837	4.8913	4.1212	3.5514	3.1158
22	8.7715	7.6446	6.7429	6.3587	6.0013	5.4099	4.9094	4.1300	3.5558	3.1180
23	8.8832	7.7184	6.7921	6.3988	6.0442	5.4321	4.9245	4.1371	3.5592	3.1197
24	8.9847	7.7843	6.8351	6.4338	6.0726	5.4510	4.9371	4.1428	3.5619	3.1210
25	9.0770	7.8431	6.8729	6.4642	6.0971	5.4669	4.9476	4.1474	3.5640	3.1220
26	9.1609	7.8957	6.9061	6.4906	6.1182	5.4804	4.9563	4.1511	3.5656	3.1227
27	9.2372	7.9426	6.9352	6.5135	6.1364	5.4919	4.9636	4.1542	3.5669	3.1233
28	9.3066	7.9844	6.9607	6.5335	6.1520	5.5016	4.9697	4.1566	3.5679	3.1237
29	9.3696	8.0218	6.9830	6.5509	6.1656	5.5098	4.9747	4.1585	3.5687	3.1240
30	9.4269	8.0552	7.0027	6.5660	6.1772	5.5168	4.9789	4.1601	3.5693	3.1242
35	9.6442	8.1755	7.0700	6.6166	6.2153	5.5386	4.9915	4.1644	3.5708	3.1248
40	9.7791	8.2438	0.1050	6.6418	6.2335	5.5482	4.9966	4.1659	3.5712	3.1250
45	9.8628	8.2825	7.1232	6.6543	6.2421	5.5523	4.9986	4.1664	3.5714	3.1250
50	9.9148	8.3045	7.1327	6.6605	6.2463	5.5541	4.9995	4.1666	3.5714	3.1250
55	9.9471	8.3170	7.1376	6.6636	6.2482	5.5549	4.9998	4.1666	3.5714	3.1250

TABLE III Values for Student's t Distribution

Degrees of Freedom	Confidence Interval / Significance Level	90% / 10%	95% / 5%	99% / 1%
1		6.314	12.706	63.657
2		2.920	4.303	9.925
3		2.353	3.182	5.841
4		2.132	2.776	4.604
5		2.015	2.571	4.032
6		1.943	2.447	3.707
7		1.895	2.365	3.499
8		1.860	2.306	3.355
9		1.833	2.262	3.250
10		1.812	2.228	3.169
11		1.796	2.201	3.106
12		1.782	2.179	3.055
13		1.771	2.160	3.012
14		1.761	2.145	2.977
15		1.753	2.131	2.947
16		1.746	2.120	2.921
17		1.740	2.110	2.898
18		1.734	2.101	2.878
19		1.729	2.093	2.861
20		1.725	2.086	2.845
21		1.721	2.080	2.831
22		1.717	2.074	2.819
23		1.714	2.069	2.807
24		1.711	2.064	2.797
25		1.708	2.060	2.787
26		1.706	2.056	2.779
27		1.703	2.052	2.771
28		1.701	2.048	2.763
29		1.699	2.045	2.756
60		1.671	2.000	2.660
120		1.658	1.980	2.617
inf.		1.645	1.960	2.576