ECONOMIES OF SCALE AND SCOPE

2

Few concepts in microeconomics, if any, are more fundamental to business strategy than economies of scale and the closely related economies of scope. Economies of scale allow some firms to achieve a cost advantage over their rivals. Economies of scale are a key determinant of market structure and entry. Even the internal organization of a firm can be affected by the importance of realizing scale economies.

We mostly think about economies of scale as a key determinant of a firm's *horizontal boundaries*, which identify the quantities and varieties of products and services that it produces. The extent of horizontal boundaries varies across industries, along with the importance of scale economies. In some industries, such as microprocessors and airframe manufacturing, economies of scale are huge and a few large firms dominate. In other industries, such as apparel design and management consulting, scale economies are minimal and small firms are the norm. Some industries, such as beer and computer software, have large market leaders (Anheuser-Busch, Microsoft), yet small firms (Boston Beer Company, Blizzard Entertainment) fill niches where scale economies are less important.

An understanding of the sources of economies of scale and scope is clearly critical for formulating and assessing competitive strategy. This chapter identifies the key sources of economies of scale and scope and provides approaches for assessing their importance.

WHERE DO ECONOMIES OF SCALE COME FROM?

Informally, when there are economies of scale and scope, "bigger is better." To facilitate identification and measurement, it is useful to define economies of scale and scope more precisely.

Definition of Economies of Scale

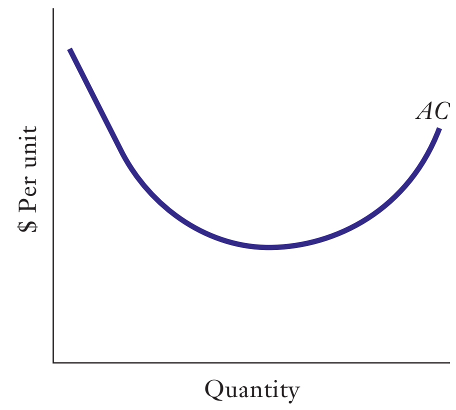
The production process for a specific good or service exhibits *economies of scale* over a range of output when average cost (i.e., cost per unit of output) declines over that range. If average cost (*AC*) declines as output increases, then the marginal cost of the last unit produced (*MC*) must be less than the average cost.[1] If average cost is increasing, then marginal cost must exceed average cost, and we say that production exhibits *diseconomies of scale*.

An *average cost curve* captures the relationship between average costs and output. Economists often depict average cost curves as U-shaped, as shown in Figure 2.1, so that average costs decline over low levels of output but increase at higher levels of output. A combination of factors may cause a firm to have U-shaped costs. A firm's average costs may decline initially as it spreads *fixed costs* over increasing output. Fixed costs are insensitive to volume; they must be expended regardless of the total output. Examples of such volume-insensitive costs are manufacturing overhead expenses, such as insurance, maintenance, and property taxes. As output increases, these costs are averaged over greater volumes, tending to drive down average costs. Firms may eventually see an upturn in average costs if they bump up against capacity constraints or if they encounter coordination or other agency problems. We will develop most of these ideas in this chapter. Coordination and agency problems are addressed in Chapters 3 and 5.

If average cost curves are U-shaped, then small and large firms would have higher costs than medium-sized firms. In reality, large firms rarely seem to be at a substantial cost disadvantage relative to smaller rivals. The noted econometrician John Johnston once examined production costs for a number of industries and determined that the corresponding cost curves were closer to L-shaped than U-shaped. Figure 2.2 depicts an L-shaped cost curve. When average cost curves are L-shaped, average costs decline up to the *minimum efficient scale (MES)* of production and all firms operating at or beyond MES have similar average costs.

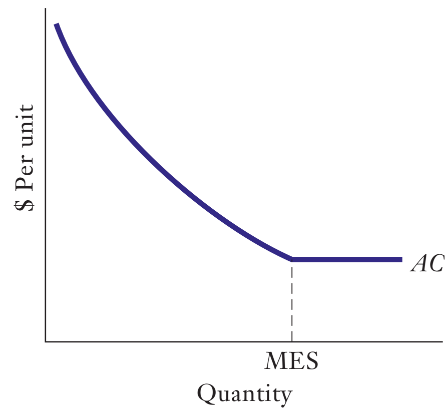
Sometimes production exhibits U-shaped average costs in the short run, as firms that try to expand output run up against capacity constraints that drive costs higher. In the long term, however, firms can expand their capacity by building new facilities.

FIGURE 2.1
A U-SHAPED AVERAGE COST CURVE



Average costs decline initially as fixed costs are spread over additional units of output. Average costs eventually rise as production runs up against capacity constraints.

Figure 2.2
An L-Shaped Average Cost Curve



When capacity does not prove to be constraining, average costs may not rise as they do in a U-shaped cost curve. Output equal to or exceeding minimum efficient scale (MES) is efficient from a cost perspective.

If each facility operates efficiently, firms can grow as large as desired without driving up average costs. This would generate the L-shaped cost curves observed by Johnston. A good example is when a cement company builds a plant in a new location or when a DVD manufacturer builds a new disc-pressing facility. We have more to say about the distinction between short- and long-run costs later in this chapter.

Definition of Economies of Scope

Economies of scale are related to economies of scope, and the two terms are sometimes used interchangeably. Economies of scale exist if the firm achieves unit-cost savings as it increases the production of a given good or service. *Economies of scope* exist if the firm achieves savings as it increases the variety of goods and services it produces. Whereas economies of scale are usually defined in terms of declining average cost functions, economies of scope are usually defined in terms of the relative total cost of producing a variety of goods and services together in one firm versus separately in two or more firms.

Because it is difficult to show scope economies graphically, we will instead introduce a simple mathematical formulation. Formally, let $TC(Q_x, Q_y)$ denote the total cost to a single firm producing Q_x units of good X and Q_y units of good Y. Then a production process exhibits scope economies if

$$TC(Q_x, Q_y) < TC(Q_x, 0) + TC(0, Q_y).$$

This formula captures the idea that it is cheaper for a single firm to produce both goods X and Y than for one firm to produce X and another to produce Y. To provide another interpretation of the definition, note that a firm's total costs are zero if it produces zero quantities of both products, so $TC(0, 0) = 0$. Then rearrange the preceding formula to read:

$$TC(Q_x, Q_y) - TC(0, Q_y) < TC(Q_x, 0)$$

This says that the incremental cost of producing Q_x units of good X, as opposed to none at all, is lower when the firm is producing a positive quantity Q_y of good Y.

The cost implications of economies of scope are shown in Table 2.1, which shows the production costs of a hypothetical manufacturer of adhesive message notes (good X) and tape (good Y). To produce tape, the firm must spend \$100 million to perfect the process of working with chemical adhesives, attaching these adhesives to cellophane, and manufacturing and packaging tape. Once this setup cost is incurred, each roll of tape can be produced at a cost of \$.20 each. Thus, we can write $TC(0, Q_y) = \$100 \text{ million} + .20Q_y$. For example, if $Q_y = 600$ million rolls of tape, total cost is \$220 million.

Now, given that the firm has made the investment in developing the know-how for manufacturing tape, much of that know-how can be applied to producing related products, such as adhesive message notes. Suppose that the additional investment needed to ramp up production of message notes, given that the up-front setup costs in tape production have already been incurred, is \$20 million. Suppose also that the cost per ream of message notes is \$.05. Then $TC(Q_x, Q_y) = \$120 \text{ million} + .05Q_x + .20Q_y$. For example, if $Q_y = 600$ million and $Q_x = 100$ million, then total cost is \$245 million. The cost to the firm of adding message notes to its production line is only \$245 million $-$ \$220 million $=$ \$25 million.

By contrast, if the firm did not produce tape, much of the up-front investment in developing the know-how for working with chemical adhesives would have to be made just to get the expertise needed to make message notes. If developing this know-how requires an investment of \$50 million, then with a per-ream cost of \$.05, $TC(Q_x, 0) = \$50 \text{ million} + .05Q_x$. Thus, if $Q_x = 100$ million, total cost would equal \$55 million. This more than doubles the additional cost to the tape manufacturer to add message notes to its production line.

This example illustrates the economic logic of exploiting economies of scope. This logic is often known as "leveraging core competencies," "competing on capabilities," or "mobilizing invisible assets."[2] In this example it makes much more sense for the tape manufacturer to diversify into the production of message notes than it would for a firm producing unrelated products, such as a prepared-food manufacturer.

Economies of scale and scope may arise at any point in the production process, from acquisition and use of raw inputs to distribution and retailing. Although business managers often cite scale and scope economies as justifications for growth activities and mergers, they do not always exist. In some cases, bigger may be worse! Thus, it is important to identify specific sources of scale economies and, if possible, measure their magnitude. The rest of this chapter shows how to do this.

TABLE 2.1
COSTS TO PRINT MESSAGE NOTES AND TAPE (IN MILLIONS OF DOLLARS)

Q_x	Q_y	$TC(Q_x, Q_y)$
100	0	\$55
0	600	\$220
100	600	\$245
200	0	\$60
0	1200	\$340
200	1200	\$370

Where Do Scale Economies Come From?

There are four major sources of scale and scope economies:

1. Indivisibilities and the spreading of fixed costs
2. Increased productivity of variable inputs (mainly having to do with specialization)
3. Inventories
4. Engineering principles associated with the "cube-square rule"

We discuss each in detail.

Indivisibilities and the Spreading of Fixed Costs

The most common source of economies of scale is the spreading of fixed costs over an ever-greater volume of output. Fixed costs arise when there are *indivisibilities* in the production process. Indivisibility simply means that an input cannot be scaled down below a certain minimum size, even when the level of output is very small.

Web-based grocery stores such as Peapod and Webvan were once thought to have unlimited growth potential, but their enthusiasts failed to appreciate the challenge of indivisibilities. Webvan once shipped groceries from its Chicago warehouse to suburbs throughout Chicagoland. To ship to a suburb such as Highland Park, Webvan required a truck, driver, and fuel. The amount that Webvan paid for these inputs was largely independent of whether it delivered to 1 household or 10. Webvan was unable to generate substantial business in Highland Park (or other Illinois communities, for that matter), so it sent its trucks virtually empty. Webvan was unable to charge enough to recoup its fixed costs and went bankrupt. Peapod faces the same problem today, but it does enough business in densely populated neighborhoods in downtown Chicago to survive.

Indivisibilities may give rise to fixed costs, and hence scale and scope economies, at several different levels: the product level, the plant level, and the multiplant level. The next few subsections discuss the link between fixed costs and economies of scale at each of these levels.

Economies of Scale Due to Spreading of Product-Specific Fixed Costs

The production of a specific product often involves fixed costs. Product-specific fixed costs may include special equipment such as the cost to manufacture a special die used to make an aircraft fuselage. Fixed costs may also include research and development expenses such as the cost of developing graphics software to facilitate development of a new video game. Fixed costs may include training expenses such as the cost of a one-week training program preceding the implementation of a total quality management initiative. Fixed costs may also include the costs necessary to set up a production process, such as the time and expense required to set up a textbook before printing it.

Even a simple production process may require substantial fixed costs. The production of an aluminum can involves only a few steps. Aluminum sheets are cut to size, formed into a rounded shape, and then punched into the familiar cylindrical can shape. A lid with an opener is then soldered on top. Though the process is simple, a single line for producing aluminum cans can cost about $50 million. If the opportunity cost of tying up funds is 10 percent, the fixed costs expressed on an annualized basis amount to about $5 million per year.[3]

The average fixed cost of producing aluminum cans falls as output increases. To quantify this effect, suppose that the peak capacity of a fully automated aluminum can plant is 500 million cans annually (or about 1 percent of the total U.S market). The average fixed cost of operating this plant at full capacity for one year is determined by dividing the annual cost ($5,000,000) by total output (500,000,000). This works out to one cent per can. On the other hand, if the plant only operates at 25 percent of capacity, for total annual production of 125 million cans, then average fixed costs equal four cents per can. The underutilized plant is operating at a three-cent cost differential per can. In a price-competitive industry like aluminum can manufacturing, such a cost differential could make the difference between profit and loss.

Economies of Scale Due to Tradeoffs among Alternative Technologies

Suppose that a firm is considering entering the can-manufacturing business but does not anticipate being able to sell more than 125 million cans annually. Is it doomed to a three-cent per can cost disadvantage? The answer depends on the nature of the alternative production technologies and the planned production output. The fully automated technology described previously may yield the greatest cost savings when used to capacity, but it may not be the best choice at lower production levels. There may be an alternative that requires less initial investment, albeit with a greater reliance on ongoing expenses. A firm choosing this "partially automated" technology may be able to enjoy fairly low average costs even if it produces only 125 million cans annually.

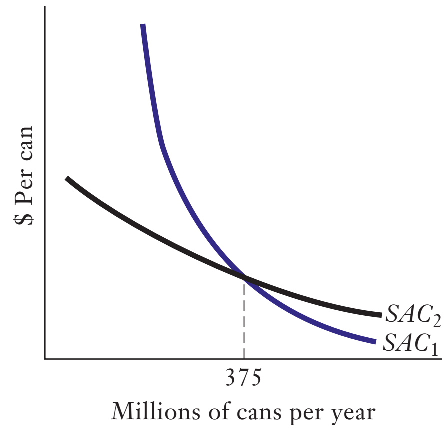
Suppose that the fixed costs of setting up a partially automated plant are $12.5 million, annualized to $1.25 million per year. The shortcoming of this plant is that it requires labor costs of one cent per can that are not needed at the fully automated plant. The cost comparison between the two plants is shown in Table 2.2.

Table 2.2 shows that while the fully automated technology has lower average total costs at high production levels, it is more costly at lower production levels. This is seen in Figure 2.3, which depicts average cost curves for both the fully and partially automated technologies. The curve labeled SAC_1 is the average cost curve for a plant that has adopted the fully automated technology; the curve labeled SAC_2 is the average cost curve for a plant that has adopted the partially automated technology. At output levels above 375 million, the fully automated technology has lower average total costs. At lower output levels, the partially automated technology is cheaper.

TABLE 2.2
COSTS OF PRODUCING ALUMINUM CANS

	500 Million Cans per Year	*125 Million Cans per Year*
Fully automated	Average fixed costs = .01 Average labor costs = .00 Average materials costs = .03 Average total costs = .04	Average fixed costs = .04 Average labor costs = .00 Average materials costs = .03 Average total costs = .07
Partially automated	Averge fixed costs = .0025 Averge labor costs = .01 Averge materials costs = .03 Averge total costs = .0425	Average fixed costs = .01 Average labor costs = .01 Average materials costs = .03 Average total costs = .05

Figure 2.3
Average Cost Curves for Can Production

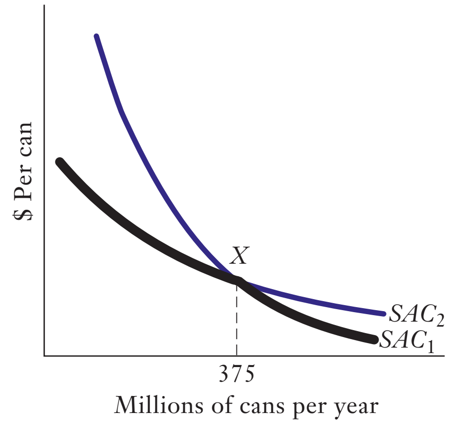


SAC_1 represents a high fixed/low variable cost technology. SAC_2 represents a low fixed cost/high variable cost technology. At low levels of output, it is cheaper to use the latter technology. At high outputs, it is cheaper to use the former.

The aluminum can example demonstrates the difference between economies of scale that arise from increased capacity utilization with a given production technology and economies of scale that arise as a firm chooses among alternative production technologies. Reductions in average costs due to increases in capacity utilization are *short-run* economies of scale in that they occur within a plant of a given size. Reductions due to adoption of a technology that has high fixed costs but lower variable costs are *long-run* economies of scale. Given time to build a plant from scratch, a firm can choose the plant that best meets its production needs, avoiding excessive fixed costs if production is expected to be low and excessive capacity costs if production is expected to be high.

Figure 2.4 illustrates the distinction between short-run and long-run economies of scale. (Chapter 1 discusses this distinction at length.) SAC_1 and SAC_2, which duplicate the cost curves in Figure 2.3, are the short-run average cost curves for the partially automated and fully automated plants, respectively. Each decreases because, as output within each plant grows, fixed costs are spread over more and more units.

Figure 2.4
Short-Run versus Long-Run Average Cost



In the long run, firms may choose their production technology as well as their output. Firms planning to produce beyond point *X* will choose the technology represented by SAC_1. Firms planning to produce less than point *X* will choose the technology represented by SAC_2. The heavy "lower envelope" of the two cost curves, which represents the lowest possible cost for each level of production, is called the *long-run average cost curve*.

EXAMPLE 2.1 Hub-and-Spoke Networks and Economies of Scope in the Airline Industry

An important example of multiplant economies of scope arises in a number of industries in which goods and services are routed to and from several markets. In these industries, which include airlines, railroads, and telecommunications, distribution is organized around "hub-and-spoke" networks. In an airline hub-and-spoke network, an airline flies passengers from a set of "spoke" cities through a central "hub," where passengers then change planes and fly from the hub to their outbound destinations. Thus, a passenger flying from, say, Omaha to Boston on United Airlines would board a United flight from Omaha to Chicago, change planes, and then fly from Chicago to Boston.

Recall that economies of scope occur when a firm producing many products has a lower average cost than a firm producing just a few products. In the airline industry, it makes economic sense to think about individual origin–destination pairs (e.g., Omaha to Boston, Chicago to Boston) as distinct products. Viewed in this way, economies of scope exist if an airline's average cost is lower the more origin–destination pairs it serves.

To understand how hub-and-spoke networks give rise to economies of *scope*, it is first necessary to explain *economies of density*. Economies of density are essentially economies of scale along a given route, that is, reductions in average cost as traffic volume on the route increases. (In the airline industry, traffic volume is measured as revenue-passenger miles [RPM], which is the number of passengers on the route multiplied by the number of miles, and average cost is the cost per revenue passenger mile.) Economies of density occur because of spreading flight-specific fixed costs (e.g., costs of the flight and cabin crew, fuel, aircraft servicing) and because of the economies of aircraft size. In the airline industry, traffic-sensitive costs (e.g., food, ticket handling) are small in relation to flight-specific fixed costs. Thus, as its traffic volume increases, an airline can fill a larger fraction of its seats on a given type of aircraft (in airline industry lingo, it increases its *load factor*—the ratio of passengers to available seats), and because the airline's total costs increase only slightly, its cost per RPM falls as it spreads the flight-specific fixed costs over more traffic volume. As traffic volume on the route gets even larger, it becomes worthwhile to substitute larger aircraft (e.g., 300-seat Boeing 767s) for smaller aircraft (e.g., 150-seat Boeing 737s). A key aspect of this substitution is that the 300-seat aircraft flown a given distance at a given load factor is less than twice as costly as the 150-seat aircraft flown the same distance at the same load factor. The reason for this is that doubling the number of seats and passengers on a plane does not require doubling the sizes of flight and cabin crews or the amount of fuel used, and that the 300-seat aircraft is less than twice as costly to build as the 150-seat aircraft, owing to the cube-square rule, which will be discussed below.

Economies of scope emerge from the interplay of economies of density and the properties of a hub-and-spoke network. To see how, consider an origin–destination pair such as Omaha to Boston. This pair has a modest amount of daily traffic. An airline serving only this route would use small planes and operate with a relatively low load factor. But now consider United's traffic on this route. United offers daily flights from Omaha to Chicago. It not only draws passengers who want to travel from Omaha to Chicago, but it would also draw passengers traveling from Omaha to all other points accessible from Chicago in the network, including Boston. By including the Omaha–Chicago route as part of a larger hub-and-spoke network, United can operate a larger airplane at higher load factors than can an airline serving only Omaha–Chicago. United benefits from economies of density to achieve a lower cost per RPM along this route. Moreover, because there will now be passengers traveling between Chicago and other spoke cities in this network, the airline's load factors on these other spokes will increase somewhat, thereby lowering the costs per RPM on these routes as well. This is precisely what is meant by economies of scope.

As more travelers take to the skies, and as smaller and more efficient jet aircraft reach the market, it is becoming possible to fly efficient nonstop flights between what were previously spoke cities. For example, Jet Blue flies nonstop from Fort Lauderdale to Long Beach; previously, this trip required flying on another carrier and changing at a hub city. This trend is reducing the economic advantages that were previously enjoyed by the major hub-and-spoke carriers.

If we trace out the lower regions of each curve, we see the long-run average cost curve. The long-run average cost curve is everywhere on or below each short-run average cost curve. This reflects the flexibility that firms have to adopt the technology that is most appropriate for their forecasted output.

Indivisibilities Are More Likely When Production Is Capital Intensive

When the costs of productive capital such as factories and assembly lines represent a significant percentage of total costs, we say that production is *capital intensive*. Much productive capital is indivisible and therefore a source of scale economies. As long as there is spare capacity, output can be expanded at little additional expense. As a result, average costs fall. Conversely, cutbacks in output may not reduce total costs by much, so average costs rise. When most production expenses go to raw materials or labor, we say that production is *materials* or *labor intensive*. Because materials and labor are divisible, they usually change in rough proportion to changes in output, with the result that average costs do not vary much with output. It follows that substantial product-specific economies of scale are more likely when production is capital intensive, and minimal product-specific economies of scale are likely when production is materials or labor intensive.

"Labor intensive technologies fail to exhibit economies of scale" is a useful rule of thumb but should not be followed slavishly. Prescription drug manufacturers rely on huge sales forces to market their drugs to physicians. But there are substantial fixed travel costs each time a sales rep visits physicians in a given specialty and market area. Drug makers can reduce average selling costs if their sales reps can promote more drugs per visit. This helps explain why drug makers are more likely to expand their offerings within specific therapeutic categories such as cancer or cardiovascular care than to diversify their offerings across categories.[4]

"The Division of Labor Is Limited by the Extent of the Market"

Economies of scale are closely related to the concept of specialization. To become specialists, individuals or firms must often make substantial investments. They will not do so unless demand justifies it; if demand is inadequate, they will not recover their costs and will be reluctant to specialize. This is the logic underlying Adam Smith's famous theorem, "The division of labor is limited by the extent of the market." (Adam Smith is the father of laissez-faire economics. His best-known work, *Wealth of Nations*, was published in 1776.) The *division of labor* refers to the

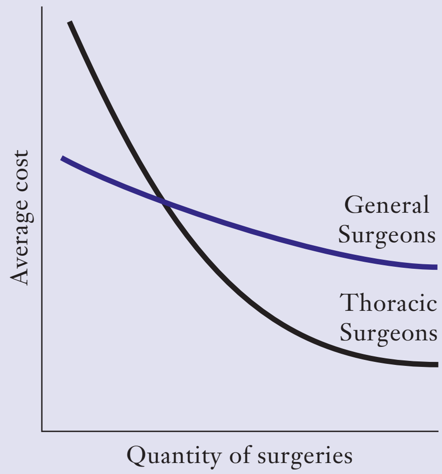
EXAMPLE 2.2 The Division of Labor in Medical Markets

An interesting application of Smith's theorem involves the specialization of medical care. Physicians may practice general medicine or specialty medicine. Generalists and specialists differ in both the amount of training they receive and the skill with which they practice. Take the case of surgery. To become general surgeons, medical school graduates spend three to four years in a surgical residency. They are then qualified to perform a wide variety of surgical procedures. Because their training is broad, general surgeons do all kinds of surgery with good, but not necessarily great, skill.

Contrast this with the training and skills of a thoracic surgeon. Thoracic surgeons specialize in the thoracic region, between the neck and the abdomen. To become a thoracic surgeon, a medical school graduate must complete a residency in general surgery and then an additional two-year residency in thoracic surgery. Figure 2.5 depicts average "cost" curves for thoracic surgery performed by a general surgeon and a thoracic surgeon. We use "cost" in quotes because it represents the full cost of care, which is lower if the surgery is successful. (Successful surgery usually implies fewer complications, shorter hospital stays, and a shorter period of recuperation.) The average cost curves are downward sloping to reflect the spreading out of the initial investments in training. The cost curve for the thoracic surgeon starts off much higher than the cost curve for the general surgeon because of the greater investment in time. However, the thoracic surgeon's cost curve eventually falls below the cost curve of a general surgeon because the thoracic surgeon will perform thoracic surgery more effectively than most general surgeons.

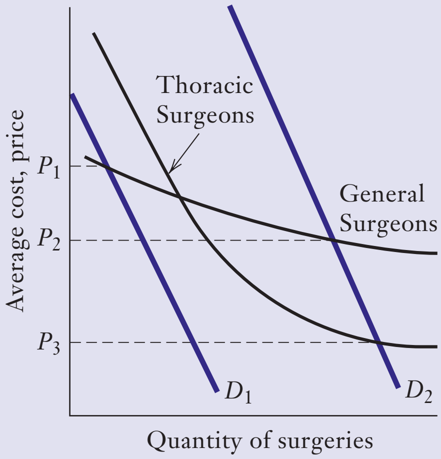
According to Smith's theorem, when the demand for thoracic surgery in a market is low, then the market will not support a specialized surgeon. Instead, thoracic surgery will be performed by a general surgeon, who may also perform other kinds of surgery. This may be seen in Figure 2.6, which superimposes demand curves over cost curves. For low levels of demand, such as at D_1, the market can support a general surgeon. A general surgeon who charges a price for thoracic surgery above P_1 can more than cover average costs. When demand is D_1, the market cannot support a thoracic surgeon.

Figure 2.5
Cost Curves for General and Thoracic Surgeons



General surgeons incur lower training costs than do thoracic surgeons but are usually less efficient in performing thoracic surgery. Thus, the general surgeon's average cost curve is below the thoracic surgeon's for low volumes (reflecting lower average fixed costs) but above the thoracic surgeon's for high volumes (reflecting higher average variable costs).

FIGURE 2.6
COST AND DEMAND FOR THORACIC SURGERY



At low demands (D_1), general surgeons may be able to cover their average costs, but dioracic surgeons may not. At high demands (D_2), thoracic surgeons may be aisle to offer lower effective prices than can general surgeons (where the effective price to the consumer includes the costs associated with ineffective surgery).

There is no price high enough to enable thoracic surgeons to recoup their costs.

When demand increases to D_2, the market can support a thoracic surgeon. A thoracic surgeon who charges a price above P_3 can cover average costs. Moreover, at prices between P_2 and P_3, the thoracic surgeon can make a profit, but the general surgeon cannot. Thus, at this high level of demand, the thoracic surgeon can drive the general surgeon from the market for thoracic surgery.

The same logic should apply to other specialized surgical and medical services. Thus, in large markets, we may expect to see a range of specialists and few or no generalists. Researchers at the RAND Corporation documented this pattern of the division of labor in medical markets.[5] They found that general practitioners are disproportionately located in smaller towns—they do not appear to fare well in larger markets, which have a wider assortment of specialists. James Baumgardner also found that physicians who practice in small towns treat a wider range of illnesses than do their big-city counterparts.[6]

specialization of productive activities, such as when a financial analyst specializes in the analysis of startup biotech companies. As suggested, this usually requires upfront investments—the analyst must do considerable research on the biotech industry before having the credibility to compete for clients. The *extent of the market* refers to the magnitude of demand for these activities, such as the demand for financial advice about start-up biotech companies. Although Smith referred mainly to specialization by individuals, his ideas apply equally well to specialization by firms.

Smith's theorem states that individuals or firms will not make specialized investments unless the market is big enough to support them. Indeed, growing prospects for startup biotech companies during the late 1980s did cause some financial analysts to specialize in this sector. One additional implication of Smith's theorem is that larger markets will support a more specialized array of activities than smaller markets can. A small town may have a pet store that caters to owners of all kinds of critters. In a big city, one can find dog groomers, saltwater aquarium boutiques, and stores that sell nothing but exotic birds.

Inventories

Economies of scale may arise when firms carry inventories. This may include traditional inventories, such as parts at an auto repair shop, and nontraditional inventories, such as customer service agents at a call center. Firms carry inventory to minimize the chances of running out of stock. A stock-out can cause lost business and cause potential customers to seek more reliable sources of supply. For a manufacturer, a stock-out for a single part may delay an entire production process. Of course, there are costs to carrying inventory, including interest on the expenses borne in producing the inventory and the risk that it will depreciate in value while waiting to be used or sold.

Inventory costs drive up the average costs of the goods that are actually sold. Suppose, for example, that a firm needs to hold inventories equal to 10 percent of its sales to maintain a tolerable level of expected stock-outs. This will increase its average cost of goods sold by as much as 10 percent. (The increase will be smaller if, at the end of the selling season, the firm can sell its inventories at some fraction of original cost.) Inventory costs are so important in some sectors, such as mass merchandising, that firms like Wal-Mart and Target are able to outcompete their rivals largely on their ability to manage inventories.

In general, inventory costs are proportional to the ratio of inventory holdings to sales. The need to carry inventories creates economies of scale because firms doing a high volume of business can usually maintain a lower ratio of inventory to sales while achieving a similar level of stock-outs. This reduces their average cost of goods sold. A full justification for this statement requires an extensive foray into the complex topic of *queuing theory*. This is well beyond the scope of this text, so instead we will offer an illustrative example.

Consider two equal-sized hospitals stocking a blood substitute that must be discarded after one month. Each hospital expects to use 20 liters per month. However, to ensure that there is only a 5 percent chance of running out, each holds 50 liters in inventory. If the blood substitute costs \$100 per liter, then each hospital has an expected average cost of \$250 per liter actually used. Suppose that the two hospitals share inventories (as they might if they merged). If one hospital runs out of the blood substitute, it can obtain it from the other. This implies that if the two hospitals maintain their present inventories of 50 liters apiece, their outage rates will be much less than 5 percent. It follows that they can maintain the desired 5 percent outage rate with lower inventories and, therefore, lower inventory holding costs. William Lynk has estimated the potential economies from hospitals sharing supplies and equipment in this way.[7] The potential saving is as large as 10 percent in some departments.

The Cube-Square Rule and the Physical Properties of Production

Economies of scale also arise because of the physical properties of processing units. An important example of this is the *cube-square rule*, well known to engineers.[8] It states that as the volume of a vessel (e.g., a tank or a pipe) increases by a given proportion (e.g., it doubles), the surface area increases by less than this proportion (e.g., it less than doubles).

What does the cube-square rule have to do with economies of scale? In many production processes, production capacity is proportional to the *volume* of the production vessel, whereas the total cost of producing at capacity is proportional to

EXAMPLE 2.3 THE ACE HARDWARE CORPORATION

When it comes to the U.S. hardware market, Home Depot and Lowes command most of the market share and most of the media attention. At the corporate level, these stores enjoy economies of scale in purchasing and marketing. At the store level, these big box stores enjoy inventory economies and offer do-it-yourselfers and professional tradesmen the benefits of one-stop shopping. It is no wonder that there are over 2,100 Home Depot and another 1,100 Lowes stores operating in all 50 states. It may therefore be surprising to know that there are far more stores operating under the Ace and True Value brand names. Ace and True Value hardware stores tend to be much smaller than Home Depot and Lowes, but that has not prevented these individually owned stores from realizing some of the same economies of scale, through membership in purchasing groups.

Ace began as a hardware wholesaler and distributor based in the Midwest. A dealer buyout in 1974 led to a national expansion. Today, the Ace cooperative is jointly owned by its 4,800 member stores spanning 70 countries. (The 8,500 members of the other large cooperative, TruServ Corporation, operate under the True Value name.) Ace purchases in bulk from more than 4,000 suppliers, including Stanley Tools, Toro (lawnmowers), and Weber (grills). This gives individual store owners access to the same distribution channels as Home Depot and Lowes at nearly comparable costs. Ace employs its own buyers, who obtain quantity discounts that are then passed on to individual stores. In fact, any cost differences between Home Depot and Ace are likely due to distribution and inventory rather than bulk purchasing discounts. Ace suppliers manufacture high-quality products that are sold under the Ace label. (Ace makes its own paint—about the only product that it makes itself.)

Ace provides its members with other benefits as well. It places national advertising and coordinates national marketing campaigns. It provides information about local marketing practices and new products. It developed and supervised the installation of electronic inventory control systems for store owners, allowing them to rapidly check on inventories and prices (this is especially helpful for products, such as lumber, which experience volatile price movements), place special orders, and communicate with other stores. Finally, Ace used its clout to push suppliers to adopt bar codes to facilitate pricing and inventory maintenance.

Individual Ace hardware stores can match the purchasing and marketing economies of national chains. At the same time, they enjoy the benefits of independence. Store owners face hard-edged market incentives that encourage leading-edge customer service. This was manifested in 2008, when *Business Week* ranked Ace in the top 10 of "Customer Service Champs" and JD Powers ranked Ace "Highest in Customer Satisfaction among Home Improvement Stores." In contrast, a 2006 University of Michigan survey ranked Home Depot last among all major U.S. retailers in customer service. Local control has other benefits. Stores can tailor their product offerings to local needs; one author's local Ace has a regional reputation as the place to shop for ladders. Stores can even procure a large fraction of their merchandise from outside the purchasing group. Ace store managers are also free to match advertised Home Depot and Lowes prices on big-ticket items such as gas grills. Finally, the smaller size of the stores allows them to locate in neighborhoods, instead of the strip malls favored by Home Depot and Lowes. As a result, most Americans live within 5 miles of an Ace store.

The cooperative concept has disadvantages, however. Absent direction from a central office, individual stores may cannibalize each other through aggressive pricing and marketing practices. Store locations are not chosen with a mind toward inventory management, though in many cases the same individual will own several local Ace locations, offering even greater scale economies (though at the expense of some store-level independence). Key decisions regarding inventories, purchasing, and marketing can be delayed owing to the democratic nature of the cooperative. And Ace lacks the standardization that assures a Home Depot customer of consistent selection and pricing at all stores.

There are many types of hardware customers—some will shop for the lowest price and can fend for themselves, while others will pay for service as long as the cost is not too dear. Thanks to the clout of purchasing cooperatives, it seems that all hardware customers can find the store that best meets their needs.

the *surface area* of the vessel. This implies that as capacity increases, the average cost of producing at capacity decreases because the ratio of surface area to volume decreases. More generally, the physical properties of production often allow firms to expand capacity without comparable increases in costs.

Oil pipelines are an excellent example of this phenomenon. The cost of transporting oil is an increasing function of the friction between the oil and the pipe. Because the friction increases as the pipe's surface area increases, transportation costs are proportional to the pipe's surface area. By contrast, the amount of oil that can be pumped through the pipe depends on its volume.[9] Thus, the average cost of a pipeline declines as desired throughput increases. Other processes that exhibit scale economies owing to the cube-square rule or related properties include warehousing (the cost of making the warehouse is largely determined by its surface area) and brewing beer (the volume of the brewing tanks determine output).

◆ ◆ ◆ ◆ ◆ Special Sources of Economies of Scale and Scope

The sources of economies of scale in the previous section related mainly to production. This section describes three special sources of economies of scale and scope having to do with areas other than production:

1. Economies of scale and scope in purchasing
2. Economies of scale and scope in advertising
3. Economies of scale and scope in research and development

Economies of Scale and Scope in Purchasing

Most of us have purchased items in bulk. Whether we are buying gallon containers of milk or six-packs of soda, the price per unit of many items falls as the amount we purchase increases. Big businesses that make large purchases from their suppliers may also obtain discounts, enabling them to enjoy a cost advantage over smaller rivals.

There is no necessary reason for big buyers to obtain bulk discounts. A supplier might not care whether it sells 100 units to a single buyer or 10 units to each of 10 different buyers. There are three possible reasons why a supplier would care:

1. It may be less costly to sell to a single buyer, for example, if each sale requires some fixed cost in writing a contract, setting up a production run, or delivering the product.
2. A bulk purchaser has more to gain from getting the best price and therefore will be more price sensitive. For example, someone purchasing hundreds of computer

printers on behalf of a university is more likely to switch vendors over small price differences than someone buying one printer for personal use.

3. The supplier may fear a costly disruption to operations—or, in the extreme case, bankruptcy—if it fails to do business with a large purchaser. The supplier may offer a discount to the large purchaser so as to assure a steady flow of business.

Small firms can take steps to offset these conditions and nullify purchasing economies. For example, small firms may form purchasing alliances that buy in bulk in order to obtain quantity discounts. This has enabled over 10,000 independent hardware stores to stay afloat as part of the Ace and True Value purchasing groups, even as Lowes and Home Depot drive down prices. To take another example, consider wholesale pricing of prescription drugs. Relatively small mail-order pharmacies often do not stock drugs for which they are unable to obtain favorable wholesale prices, whereas some pharmacy chains stock and sell drugs with little regard to their wholesale prices. As a result, mail-order pharmacies often obtain more favorable pricing terms from drug manufacturers.

Economies of Scale and Scope in Advertising

The advertising cost per consumer of a product may be expressed by the following formula:

$$\frac{\text{Cost of sending a message}}{\begin{array}{c}\text{Number of potential consumers}\\ \text{receiving the message}\end{array}} \div \frac{\begin{array}{c}\text{Number of actual consumers}\\ \text{as a result of message}\end{array}}{\begin{array}{c}\text{Number of potential consumers}\\ \text{receiving the message}\end{array}}$$

Larger firms may enjoy lower advertising costs per consumer either because they have lower costs of sending messages per potential consumer (the first term) or because they have higher advertising *reach* (the second term).

Costs of Sending Messages per Potential Consumer

There are important fixed costs associated with placing an ad, including preparation of the ad and negotiation with the publisher or broadcaster. If ad preparation costs and the costs of negotiating a single national and local advertising "buy" are about the same, the national advertiser will have a lower cost per potential consumer because these fixed costs get spread over a larger base of potential consumers.

To illustrate, suppose that Anheuser-Busch places an ad in *USA Today* and pays Gannett (the publisher of *USA Today*) \$10 per thousand papers sold to run this ad. Because *USA Today* has a daily circulation of about 2 million, the direct costs of this ad to Anheuser-Busch would be \$10 × (2,000,000/1000), or \$20,000. The same day, Hudepohl, a local brewery in Cincinnati, Ohio, places an ad in the *Cincinnati Enquirer* (the local paper) and, let's say, pays the same rate of \$10 per thousand papers sold. The *Enquirer* has a daily circulation of about 200,000, so the direct cost to Hudepohl would be \$10 × (200,000/1,000), or \$2,000. Finally, suppose that for both companies the cost of preparing the ad is \$4,000.

Let us now look at the advertising cost per potential consumer for Anheuser-Busch and Hudepohl:

- Anheuser-Busch advertising cost per potential consumer = (\$20,000 + \$4,000)/2,000,000 = \$.012 per potential consumer, or \$12 per 1,000 potential consumers.

- Hudepohl advertising cost per potential consumer = ($2,000 + $4,000)/200,000 = $.030 per potential consumer, or $30 per 1,000 potential consumers.

This example illustrates the approximate difference in the cost per potential consumer between national and local advertising.

The logic underlying this example illustrates why national firms, such as McDonald's, enjoy an advertising cost advantage over their local counterparts, such as Gold Coast Dogs (a chain of hot dog restaurants in Chicago).

Advertising Reach and Umbrella Branding

Even when two firms have national presences, the larger one may still enjoy an advantage. Suppose that Wendy's and McDonald's both place advertisements on rival television networks to air at the same time. The ads are seen by audiences of equal sizes and cost the same to place. Both ads are equally persuasive—20,000 viewers of the McDonald's ad have an urge to visit McDonald's; 20,000 viewers of the Wendy's ad are motivated to visit Wendy's. Despite these similarities, the cost per effective message is much lower for McDonald's. The reason is that there are about three times as many McDonald's in the United States as there are Wendy's. Almost all of the 20,000 viewers craving McDonald's can find one nearby, but many of the 20,000 who crave Wendy's cannot.

The effectiveness of a firm's ad may also be higher if that firm offers a broad product line under a single brand name. For example, an advertisement for a Samsung flat-screen plasma television may encourage customers to consider other products made by Samsung, such as DVD players. This is known as *umbrella branding*. Umbrella branding is effective when consumers use the information in an advertisement about one product to make inferences about other products with the same brand name, thereby reducing advertising costs per effective image. When Samsung advertises its plasma television, consumers may infer that Samsung is on the cutting edge of technology and therefore that its other high-tech products are also good.

Umbrella branding does not always work. Drug companies have successfully marketed individual brand-name drugs (e.g., erectile dysfunction drugs such as Levitra), but their corporate brands do not have much influence with consumers (few Levitra consumers know that the product is made by GlaxoSmithKline). Sometimes firms prefer to keep brand identities separate—Toyota launched the Lexus nameplate to avoid "tarring" its luxury cars with a mass-market reputation. Some brands are not meant to share the same image. In the 1970s, British conglomerate EMI's medical division introduced its new CT scanner medical diagnostic equipment. Potential buyers were shocked when EMI's music division signed a contract with punk pioneers the Sex Pistols. The corporate parent quickly dumped the Pistols (prompting Johnny Rotten to write a song entitled "EMI" about the incident). EMI was unable to sign any significant "new wave" performers for many years thereafter.

Economies of Scale in Research and Development

R&D expenditures exceed 5 percent of total sales revenues at many companies, including Intel, Microsoft, GlaxoSmithKline, and GE. The nature of engineering and scientific research implies that there is a minimum feasible size to an R&D project as well as an R&D department. For example, researchers at Tufts University have carefully measured the costs of developing new pharmaceutical products for the U.S. market.[10] They estimate that drug companies must spend upwards of $500 million to

EXAMPLE 2.4 The Pharmaceutical Merger Wave

Beginning in the 1990s, pharmaceutical companies faced an unprecedented strategic challenge. The growth of managed care in the United States and the tightening of government health care budgets in other nations forced manufacturers to lower prices on many drugs. Traditional research pipelines began to dry up, while the advent of biotechnology promised new avenues for drug discovery coupled with new sources of competition. In response to these pressures, the pharmaceutical industry underwent a remarkable wave of consolidation, with the total value of merger and acquisition activity exceeding $500 billion. As a result, the combined market shares of the 10 largest firms have grown from 20 percent to more than 50 percent. Using almost any yardstick, we can view Glaxo's 2000 acquisition of SmithKline Beecham and Pfizer's 2003 acquisition of Pharmacia as among the largest in business history.

Industry analysts point out three potential rationales for consolidation. One cynical view is that executives at struggling pharmaceutical companies are buying the research pipelines of more successful rivals merely to save their jobs. We explore such managerial rationales for merger in Chapter 6.

Another potential rationale is to make more efficient use of sales personnel. Many pharmaceutical firms spend more money on sales than they do on R&D. Although pharmaceutical "direct to consumer" advertising has received a lot of attention lately, drug makers spend much more money on traditional advertising in medical journals and especially on "detailing." Detailing is when sales personnel visit doctors and hospitals to describe the benefits of new drugs and share data on efficacy and side effects. Detailers spend most of their time on the road, creating an obvious opportunity for scale economies. A detailer who can offer several cardiovascular drugs to a cardiologist will have a much higher ratio of selling time to traveling time. Why have two detailers from two companies visiting the same cardiologist when one will do?

Perhaps the most common explanation offered for the merger wave is to exploit economies of scale in R&D. As we discuss in the accompanying text, there are conflicting theories as to whether bigger firms will be more innovative or will innovate at lower cost. The theoretical considerations apply especially well in pharmaceutical R&D, and those in the industry who bank on achieving greater research effectiveness through scale economies in R&D may not have solid footing.

Recent research by Patricia Danzon, Andrew Epstein, and Sean Nicholson examines some of these potential scale economies.[12] Looking at financial and sales data for over 200 pharmaceutical companies spanning the time period 1988–2000, they found that acquirers tended to have older drug portfolios, lending some support to the cynical explanation for acquisitions. In contrast, targets had average to possibly slightly younger portfolios. Combined sales after the merger seemed to be slightly below premerger sales levels, which may reflect the weak portfolios of the acquirers. Addressing scale economies, they find that two years after a merger, the number of employees had fallen by about 6 percent. This finding is consistent with economies of scale in sales. R&D spending does not change postmerger. Because of the long lag between R&D spending and new products reaching the market, it is too soon to tell if R&D productivity has increased.

successfully develop a new drug. This is a substantial indivisible investment, implying that average fixed costs will decline rapidly as sales of a particular drug increase.

R&D may also entail economies of scope if ideas developed in one research project create positive spillovers to another project. Using detailed R&D data, Rebecca Henderson and Iain Cockburn looked for evidence of scope economies in pharmaceutical firms.[11] Using the the number of patents per dollar of R&D as a measure of productivity, they found that an average firm with 19 research programs was 4.5 percent more productive than a firm with 17 programs. The overall evidence linking size and innovativeness is ambiguous, however. Large firms may benefit from spillovers, but smaller firms may have greater incentives to innovate. Moreover, smaller firms may take a variety of independent approaches to tackling research problems, whereas a large firm may pursue a narrow agenda more aggressively. Depending on the nature of the science, either approach may be quicker to yield fruitful outcomes. We elaborate on these ideas in Chapter 14. Suffice it to say that economic theory and empirical evidence is ambiguous about whether big firms are more innovative than small firms.

Complementarities and Strategic Fit

Economists usually use the concept of scope economies to describe the synergies enjoyed by a firm that produces an array of complementary products and services. Paul Milgrom and John Roberts coined the term *complementarities* to describe synergies among organizational practices.[13] Practices display complementarities when the benefits of introducing one practice are enhanced by the presence of others. For example, Southwest Airlines strives for the fastest turnaround of any airline, often landing a plane and departing within 30 minutes. To do this, Southwest uses several complementary practices. It does not cater its flights. It uses a single type of plane (Boeing 737), thereby simplifying baggage handling, refueling, and maintenance procedures. It does not fly into congested airports. Each of these practices makes the others more effective by eliminating potential bottlenecks. Thus, the reduction in maintenance time afforded by the use of a single type of plane would be wasted if Southwest took the time to cater meals.

The concept of complementarities is better known in the strategy literature as *strategic fit*. Harvard Business School Professor Michael Porter has argued that strategic fit among processes is essential to firms seeking a long-term competitive advantage over their rivals. Through strategic fit, the "whole" of a firm's strategy exceeds the "sum of the parts" of its organizational processes. Moreover, it is difficult for other firms to copy the strategy because they would have to successfully copy each individual process. For example, United Airlines could switch to a single type of plane, or stop onboard catering, but unless it moved out of its congested Chicago hub, it could not hope to match Southwest's operational efficiencies.

The power of strategic fit can been seen by a simple mathematical exercise. Suppose that a firm like Southwest has successfully implemented ten different organizational practices. Its rivals can observe these practices and try to emulate them. But suppose that the probability of successfully copying any one practice is only .80, either because Southwest possesses unique skills or, what is more likely, the history of the competition restricts what they can do. In this case, the probability of copying all ten practices equals $.80^{10} = .11$, or 11 percent. Not only are Southwest's rivals unlikely to copy all ten practices, complementarity among the ten practices implies that there is a substantial disadvantage to firms that can copy even eight or nine of them.

◆ ◆ ◆ ◆ ◆ Sources of Diseconomies of Scale

Given the attention we have paid to scale and scope economies, one might expect some colossal "megafirm" to dominate production across all industries. Antitrust laws may place some limits to firm growth. More likely, though, firms understand that there are limits to economies of scale, so that beyond a certain size, bigger is no longer better and may even be worse. Diseconomies of scale may arise for a number of reasons; here are some of the most important.

Labor Costs and Firm Size

Larger firms generally pay higher wages and provide greater benefits. Even if one controls for other determinants of wages such as work experience and job type, a wage gap of 10 percent or more between large and small firms is not unusual. Labor economists offer several possible reasons for the wage gap. Large firms are more likely to be unionized than small firms. Workers in smaller firms may enjoy their work more than workers in large firms, forcing large firms to pay a *compensating differential* to attract workers. Large firms may need to draw workers from greater distances, forcing them to pay a compensating differential to offset transportation costs. Some economists speculate that the wage premium reflects hard-to-measure aspects of worker quality, such as their skill and experience in capital-intensive production processes. According to this view, size itself does not handicap larger firms. Instead, large firms are merely paying a premium to workers with unique and highly valued skills.

Two factors work in favor of larger firms. First, worker turnover at larger firms is generally lower. Since it can cost thousands of dollars to recruit and train new employees, this may offset some of the added costs due to higher wages. Second, large firms may be more attractive to highly qualified, upwardly mobile workers who want to move up the corporate ladder without changing employers.

Spreading Specialized Resources Too Thin

Many talented individuals believe that having achieved success in one venue, they can duplicate it elsewhere. Sometimes this is sheer hubris, such as when Donald Trump felt that lending his name, but not his personal attention, to Atlantic City casinos would be enough to assure their success. (Trump Hotel & Casino Resorts filed for bankruptcy in November 2004.) Others fail because they lack the skills necessary to translate their success to a new situation, such as when investment wiz Edward Lampert purchased and assumed management control of Sears. Some individuals simply spread themselves too thin. There are many stories of successful chef/owners opening a second and third restaurant, only to see the performance of all of their restaurants decline.

The same lessons also apply to specialized capital inputs, such as computers, tools and dies, or assembly lines. If a specialized input is a source of advantage for a firm, and that firm attempts to expand its operations without duplicating the input, the expansion may overburden the specialized input.

"Conflicting Out"

Professional services firms in marketing, accounting, consulting, and law face another source of diseconomies of scale—*conflicting out*. When a potential client approaches a professional services firm with new business, it may be concerned about whether the firm is already doing business with one or more of its competitors and the resulting potential conflict of interest. Faced with these concerns, the client may take its business elsewhere. The professional services firm will have been conflicted out.

The possibilities for conflict place a natural limit on the market share that any one professional services firm can achieve. For example, the marketing firm Chiat/Day lost its Coke account in 1995 when it was acquired by the Omnicom Group, which already owned BBD&O, the main ad agency for Pepsi.

Incentive and Coordination Effects

Chapter 5 describes a number of incentive effects that make it difficult for firms to expand their vertical boundaries. The same problems emerge as firms expand their horizontal boundaries. In larger firms, compensation is much less likely to be tied to the worker's contribution toward firm profit. Larger firms may also have a more difficult time monitoring and communicating with workers, further leading to difficulties in promoting effective worker performance.

EXAMPLE 2.5 The AOL Time Warner Merger and Economies of Scope[14]

When AOL and Time Warner announced their merger in January 2000, Wall Street analysts rejoiced at the possibilities because potential synergies abounded. At the tail end of the technology boom, the merger was served up as the ultimate marriage between the Old Economy and the New Economy.

Less than one year after the merger hoopla, it was already obvious that AOL Time Warner could not immediately deliver on promises of 12 to 15 percent annual growth. The failure to meet projections was partially attributed to the failing economy. Two years later, AOL was still struggling to meet expectations, and CEO Stephen Case had been forced to resign. Today, the merger is regarded as a poster child for the unrealistic expectations of the dot.com era.

By joining forces, AOL Time Warner had hoped to become the leader among companies striving for convergence among media, technology, and communications. Entering the merger, AOL had a dominant online presence. The leading Internet service provider in the world, AOL had 29 million subscribers, who accounted for 33 percent of all time spent on the Internet. This figure was greater than that of any other service or content provider, as second and third place MSN and Yahoo! each accounted for 7 percent.[15] In fact, AOL had already become something of a pop icon, and its robotically voiced e-mail delivery tagline "You've got mail" was featured in the hit movie of the same name. Even so, the growing popularity of broadband was slowly cutting AOL's business, and the effort to integrate with Time Warner may have slowed AOL's move to broadband. As broadband competition grew, AOL contemplated moving its business model from one driven by subscriber revenue to one driven by free access to content combined with advertising revenue.

Before the merger, Time Warner had also exhibited a strong off-line franchise in multiple distribution channels. At the time, it was the nation's second-largest cable operator. Time Warner cable networks like CNN, HBO, TBS, and TNT accounted for 25 percent of cable viewers. Its publishing arm included 35 magazines with a circulation of 200 million. Time Warner also owned the rights to multiple hit TV shows like *ER* and *Friends*. Lastly, the company also owned top-flight movie studios in Warner Brothers and New Line Cinema.

In the years immediately after the merger, the merged company hoped to bundle top-quality content supplied by Time Warner with its service provider franchise. The company also hoped to cross-promote its various media content, activities, events, and releases across the Internet, cable, television, publication, and film. Although these ideas sounded good on paper, the reality was that many of the expected scope economies were illusory. Cross-promotional activities may boost demand, but they come at the expense of selling promotional space to outside entities. Consumers preferred the freedom to download whatever content they wished (and the freedom to do so offered by broadband providers) rather than be force-fed whatever their service provider offered. The popularity of music downloads was an especially big blow to AOL Time Warner's dream of dominating through content.

While synergies proved to be few and far between, AOL Time Warner struggled with traditional merger issues: Whose interests would the CEO represent? Could the two cultures coexist? Would independent firms prove more flexible, beating AOL Time Warner to new market opportunities? By 2002, it was apparent that the merger was failing and that AOL was proving to be a drag on its Old Economy partner. Stephen Case, AOL's billionaire founder who had taken over as CEO of the merged entity, resigned and was replaced by Richard Parsons, who had previously overseen Time Warner's content businesses. The company may not have downsized since then, but the name has—it is now known once again as Time Warner.

The Learning Curve

Medical students are encouraged to learn by the axiom "See one, do one, teach one." This axiom grossly understates the importance of experience in producing skilled physicians—one surgery is not enough! Experience is an important determinant of ability in many professions, and it is just as important for firms. The importance of experience is conveyed by the idea of the learning curve.

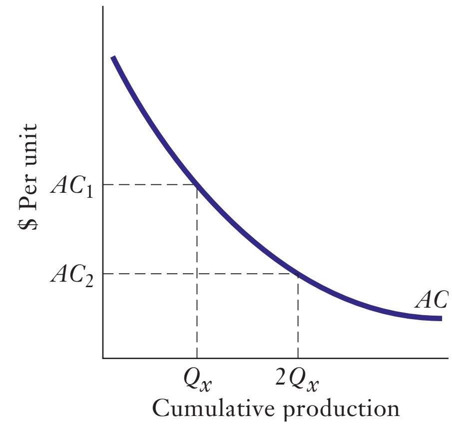
The Concept of the Learning Curve

Economies of scale refer to the advantages that flow from producing a larger output at a given point in time. The *learning curve* (or experience curve) refers to advantages that flow from accumulating experience and know-how. It is easy to find examples of learning. A manufacturer can learn the appropriate tolerances for producing a key system component. A retailer can learn about community tastes. An accounting firm can learn the idiosyncrasies of its clients' inventory management. The benefits of learning manifest themselves in lower costs, higher quality, and more effective pricing and marketing.

The magnitude of learning benefits is often expressed in terms of a *slope*. The slope for a given production process is calculated by examining how far average costs decline as cumulative production output doubles. It is important to use cumulative output rather than output during a given time period to distinguish between learning effects and other scale effects. As shown in Figure 2.7, suppose that a firm has cumulative output of Q_x with average cost of production of AC_1. Suppose next that the firm's cumulative output doubles to $2Q_x$ with average cost of AC_2. Then the slope equals AC_2/AC_1.

Slopes have been estimated for hundreds of products. The median slope appears to be about .80, implying that for the typical firm, doubling cumulative output

FIGURE 2.7
THE LEARNING CURVE



When there is learning, average costs fall with cumulative production. Here, as cumulative production increases from Q_x to $2Q_x$, the average cost of a batch of output falls from AC_1 to AC_2.

reduces unit costs by about 20 percent. Slopes vary from firm to firm and industry to industry, however, so that the actual slope enjoyed by any one firm for any given production process generally falls between .70 and .90 and may be as low as .6 or as high as 1.0 (e.g., no learning). Note that estimated slopes usually represent averages over a range of outputs and do not indicate whether or when learning economies are fully exploited.

While most studies of the learning curve focus on costs, some studies have documented the effects of learning on quality. Example 2.6 discusses a recent study of learning in medicine, where experience can literally be a matter of life and death.

Expanding Output to Obtain a Cost Advantage

When firms benefit from learning, they may want to ramp up production well past the point where the additional revenues offset the added costs. This strategy makes intuitive sense, because it allows the firm to move down the learning curve and realize lower costs in the future. Though it might seem to violate the cardinal rule of equating marginal revenue to marginal cost (see the Chapter 1), the strategy is in fact completely consistent with this rule if one properly construes the cost of current production in the presence of learning. To see why this is so, consider the following example:

Suppose that a manufacturer of DRAM chips has cumulative production of 10,000 chips. The cost to manufacture one additional chip is $2.50. The firm believes that once it has produced 100,000 chips, its unit costs will fall to $2.00, with no further learning benefits. The company has orders to produce an additional 200,000 chips when it unexpectedly receives an offer to bid on an order for 10,000 chips to be filled immediately. The firm must determine the lowest price it would be willing to accept for this order.

Assuming that filling the new order does not create delays that jeopardize other business, the firm need only compute the marginal cost of producing additional chips. If the firm myopically ignores learning effects, it would accept the order only if the price was at least $2.50 per chip. This would be wrong—the true marginal cost is not $2.50.

82476 c02.3d GGS 3/17/09 15:15

To determine the true marginal cost, the chip maker must consider how its accumulated experience will affect future costs. Before it received the new order, the chip maker had planned to produce 200,000 chips. The first 100,000 would cost $2.50 per chip, and the remaining 100,000 would cost $2.00 per chip, for a total of $450,000 for 200,000 chips. If the firm takes the new order, then the cost of producing the next 200,000 chips is only $445,000 (90,000 chips @ $2.50 + 110,000 chips @ $2.00).

By filling the new order, the DRAM manufacturer reduces its future production costs by $5,000. In effect, the incremental cost of filling the additional order is only $20,000, which is the current costs of $25,000 less the $5,000 future cost savings. Thus, the true marginal cost per chip is $2.00. The firm should be willing to accept any price over this amount, even though a price between $2.00 and $2.50 per chip does not cover current production costs.

In general, when a firm enjoys the benefits of a learning curve, the marginal cost of increasing current production is the expected marginal cost of the last unit of production the firm expects to sell. (This formula is complicated somewhat by discounting of future costs.) This implies that learning firms should be willing to accept short-run prices that are below short-run costs. They may even earn negative accounting profits in the short run but will prosper even more in the long run.

Managers who are rewarded on the basis of short-run profits may be reluctant to exploit the benefits of the learning curve. Firms could solve this problem by directly accounting for learning curve benefits when assessing profits and losses. Few firms that aggressively move down the learning curve have accounting systems that properly measure marginal costs, however, and instead rely on direct growth incentives while placing less emphasis on profits.

The Boston Consulting Group Growth/Share Paradigm

Beginning 30 years ago, the Boston Consulting Group (BCG) has been preaching aggressive growth strategies as a way of exploiting the learning curve. Figure 2.8 depicts a typical BCG *growth/share matrix*. The matrix distinguishes a firm's product lines on two dimensions: growth of the market in which the product is situated, and

Figure 2.8
The BCG Growth/Share Matrix

The growth/share matrix divides products into four categories according to their potential for growth and relative market share. Some strategists recommended that firms use the profits earned from cash cows to ramp up production of rising stars and problem children. As the latter products move down their learning curves, they become cash cows in the next investment cycle.

		Relative Market Share	
		High	Low
Relative Market Growth	High	Rising star	Problem child
	Low	Cash cow	Dog

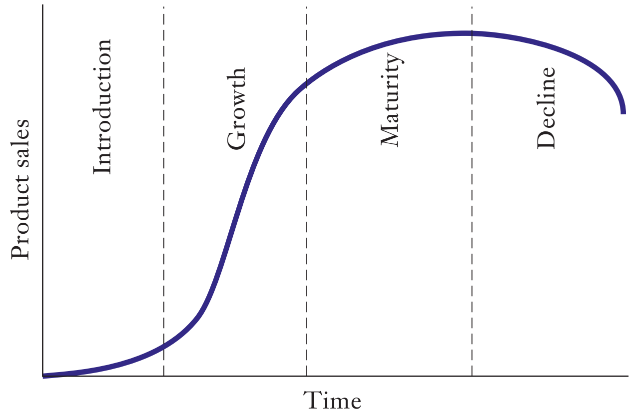
the product's market share relative to the share of its next-largest competitors. A product line was classified into one of four categories. A *rising star* is a product in a growing market with a high relative share. A *cash cow* is a product in a stable or declining market with a high relative share. A *problem child* is a product in a growing market with a low relative share. A *dog* is a product in a stable or declining market with a low relative share.

The BCG strategy for successfully managing a portfolio of products was based on taking advantage of learning curves and the *product life cycle*.[16] According to this product life-cycle model, demand for the product is initially low just after it is introduced. The product then enters a phase in which demand grows rapidly. As demand becomes increasingly driven by replacement sales rather than sales to new customers, demand growth levels off, and the product reaches its maturity stage. Finally, as superior substitute products eventually emerge, demand for the product will begin to decline. This characteristic life cycle is shown in Figure 2.9.

BCG felt that its clients should increase production in the early stages of the product's life cycle to secure learning economies. Firms could use profits from cash cow products to fund increased production of problem child and rising star products. Learning economies would cement the advantages of rising stars while enabling some problem children to become more competitive. As their markets matured and demand slackened, these products would then become cash cows to support learning strategies in new emerging markets.

BCG deserves credit for recognizing the strategic importance of learning curves. However, it would be a mistake to apply the BCG framework without considering its underlying principles. As we have discussed, learning curves are by no means ubiquitous or uniform where they do occur. At the same time, product life cycles are easier to identify after they have been completed than during the planning process. Many products ranging from nylon to the Segway personal transporter that were forecast to have tremendous potential for growth did not meet expectations. Finally, the role of the firm as "banker"—using retained earnings to fund new ventures—is questionable in an age when other sources of capital are so easily available. We pursue these ideas in more detail in Chapter 6, where we discuss diversification.

FIGURE 2.9
THE PRODUCT LIFE CYCLE



Product demand is thought to move through four stages. When the product is first introduced, sales and growth are low. Product demand then grows rapidly, but sales level off, and the industry enters a maturity phase. Eventually, demand declines as other superior products or technologies supplant it. It can be difficult to predict when each stage will begin.

Learning and Organization

While it is common to talk about organizational learning, the obvious fact is that individuals learn. Complex tasks, such as the design and production of statistical software, offer especially good opportunities for individuals to learn on their own and from their coworkers. Although individuals do the learning, firms can take steps to improve learning and the retention of knowledge in the organization. Firms can facilitate the adoption and use of newly learned ideas by encouraging the sharing of information, establishing work rules that include the new ideas, and reducing turnover. Lanier Benkard argues that labor policies at Lockheed prevented the airframe manufacturer from fully exploiting learning opportunities in the production of the L-1011 TriStar.[17] Its union contract required Lockheed to promote experienced line workers to management, while simultaneously upgrading workers at lower levels. This produced a domino effect whereby as many as 10 workers changed jobs when one was moved to a management position. As a result, workers were forced to relearn tasks that their higher-ranking coworkers had already mastered. Benkard estimates that this and related policies reduced labor productivity at Lockheed by as much as 40 to 50 percent annually.

While codifying work rules and reducing job turnover facilitates retention of knowledge, it may stifle creativity. At the same time, there are instances where worker-specific learning is too complex to transmit across the firm. Examples include many professional services, in which individual knowledge of how to combine skills in functional areas with specific and detailed knowledge of particular clients or markets may give individuals advantages that they cannot easily pass along to others. Clearly, an important skill of managers is to find the correct balance between stability and change so as to maximize the benefits of learning.

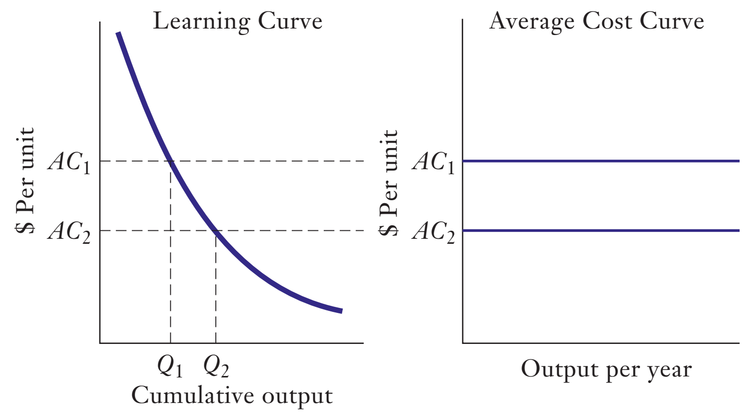
Managers should also draw a distinction between firm-specific and task-specific learning. If learning is task-specific rather than firm-specific, then workers who acquire skill through learning may be able to shop around their talents and capture the value for themselves in the form of higher wages. When learning is firm-specific, worker knowledge is tied to their current employment, and the firm will not have to raise wages as the workers become more productive. Managers should encourage firm-specific learning but must usually rely on their judgment to determine if learning is firm- or task-specific.

The Learning Curve versus Economies of Scale

Economies of learning differ from economies of scale. Economies of scale refer to the ability to perform an activity at a lower unit cost when it is performed on a larger scale at a particular point in time. Learning economies refer to reductions in unit costs due to accumulating experience over time. Economies of scale may be substantial even when learning economies are minimal. This is likely to be the case in simple capital-intensive activities, such as two-piece aluminum can manufacturing. Similarly, learning economies may be substantial even when economies of scale are minimal. This is likely to be the case in complex labor-intensive activities, such as the practice of antitrust law.

Figure 2.10 illustrates how one can have learning economies without economies of scale. The left side of the figure shows a typical learning curve, with average costs declining with cumulative experience. The right side shows two average cost curves, for different experience levels. Both average cost curves are perfectly flat, indicating

FIGURE 2.10
LEARNING ECONOMIES WHEN SCALE ECONOMIES ARE ABSENT



It is not necessary to have economies of scale to realize learning economies. The production process depicted here shows constant returns to scale, as evidenced by the flat average cost curves, which show output *within a given year*. The level of average cost falls with cumulative experience *across several years*, however, as shown by the learning curve.

that there are no economies of scale. Suppose that the firm under consideration enters a given year of production with cumulative experience of Q_1. According to the learning curve, this gives it an average cost level of AC_1. This remains constant regardless of current output because of constant returns to scale. Entering the next year of production, the firm has cumulative output of Q_2. Its experiences in the previous year enable the firm to revamp its production techniques. In thus moving down the learning curve, it can enjoy an average cost level of AC_2 in the next year of production.

Managers who do not correctly distinguish between economies of scale and learning may draw incorrect inferences about the benefits of size in a market. For example, if a large firm has lower unit costs because of economies of scale, then

EXAMPLE 2.6 LEARNING BY DOING IN MEDICINE

Learning curves are usually estimated for costs—as firms accumulate experience, the cost of production usually falls. But learning manifests itself in other ways, perhaps none as vital as in medicine, where learning can literally be a matter of life and death.

Researchers have long noted that high-volume providers seem to obtain better outcomes for their patients. This volume/outcome relationship appears dramatically in the so-called January/July effect. This is the well-documented fact that mortality rates at teaching hospitals spike in early January and July. One might explain the January spike as the after-effect of New Year's Eve revelry, but that won't explain July. The real reason is that medical residents usually change their specialty rotations in January and July. Hospital patients during these time periods are therefore being treated by doctors who may have no experience treating their particular ailments. Many other studies document the problems of newly minted physicians.

But the volume/outcome relationship also applies to established physicians. Back in the 1970s, this was taken as *prima facie* evidence of a learning curve. But there is another plausible explanation for the relationship—perhaps high-quality physicians receive more referrals. If so, then outcomes drive volume, not vice versa. This might not matter to patients who would clearly be served by visiting a high-volume provider regardless of how this chicken/egg question was resolved, but it would matter to policy makers, who have often proposed limiting the number of specialists in certain fields on the grounds that entry dilutes learning.

There is a statistical methodology that can be used to sort out causality. The technique, commonly known as instrumental variables regression, requires identifying some phenomenon that affects only one side of the causality puzzle. In this case, the phenomenon would have to affect volume but not outcomes. Statistical analysis could then be used to unambiguously assess whether higher volumes really do lead to better outcomes.

In a recent study, Subramaniam Ramanarayanan used instrumental variables regression to study the learning curve.[18] He studied cardiac surgery, where mortality rates for physicians can vary from below 2 percent to above 10 percent. As an instrument, Ramanarayanan chose the retirement of a geographically proximate heart surgeon. When a surgeon retires, volumes of other surgeons can increase by 20 patients or more annually. Retirement is a good instrument because it affects volumes but does not otherwise affect outcomes. Ramanarayanan found that surgeons who treat more patients after the retirement of a colleague enjoy better outcomes. Each additional surgical procedure reduces the probability of patient mortality by 0.14 percent. This reduction is enjoyed by all of the surgeon's patients. Ramanarayanan's study offers compelling evidence that surgeons need to maintain high volumes to be at their best.

any cutbacks in production volume will raise unit costs. If the lower unit costs are the result of learning, the firm may be able to cut current volume without necessarily raising its unit costs. To take another example, if a firm enjoys a cost advantage due to a capital-intensive production process and resultant scale economies, then it may be less concerned about labor turnover than a competitor that enjoys low costs due to learning a complex labor-intensive production process.

Chapter Summary

- A production process exhibits economies of scale if the average cost per unit of output falls as the volume of output increases. A production process exhibits economies of scope if the total cost of producing two different products or services is lower when they are produced by a single firm instead of two separate firms.
- An important source of economies of scale and scope is the spreading of indivisible fixed costs. Fixed costs do not vary as the level of production varies.
- In general, capital-intensive production processes are more likely to display economies of scale and scope than are labor- or materials-intensive processes.
- In some industries, such as food retailing, firms may make expenditures to create scale economies that previously did not exist, such as expenditures to create and reinforce brand image.
- There are economies of scale in inventory management, so that processes with large volumes need to carry less inventory on a percentage-of-output basis than similar processes with small volumes.

- The physical property known as the cube-square rule confers scale economies on processes, such as warehousing, where costs are related to the geometric volume of the production "vessel."
- There are often economies of scale associated with marketing expense, research and development, and purchasing. Large-scale marketing efforts often have lower costs per message received than do smaller-scale efforts. The costs of large research ventures may be spread over greater output, although big size may be inimical to innovation. Small firms may obtain purchasing discounts comparable to those obtained by large firms by forming purchasing groups.
- Sometimes large size can create inefficiencies. These may result from higher labor costs, agency problems, or dilution of specialized resources.
- Individuals and firms often improve their production processes with experience. This is known as learning. In processes with substantial learning benefits, firms that can accumulate and protect the knowledge gained by experience can achieve superior cost and quality positions in the market.
- (Appendix) Regression analysis that compares costs and outputs of firms of varying sizes and experience may be used to identify scale economies and the learning curve.

Questions

1. A firm produces two products, X and Y. The production technology displays the following costs, where $C(i, j)$ represents the cost of producing i units of X and j units of Y:

$$C(0,50) = 100 \qquad C(5,0) = 150$$
$$C(0,100) = 210 \qquad C(10,0) = 320$$
$$C(5,50) = 240 \qquad C(10,100) = 500$$

Does this production technology display economies of scale? Of scope?

2. Economies of scale are usually associated with the spreading of fixed costs, such as when a manufacturer builds a factory. But the spreading of fixed costs is also important for economies of scale associated with marketing, R&D, and purchasing. Explain.
3. What is the difference between economies of scale and learning economies? If a larger firm has lower average costs, can you conclude that it benefits from economies of scale? Would a small firm necessarily enjoy the same cost position if it were to duplicate the size of its larger rival?
4. A firm contemplating entering the breakfast cereal market would need to invest \$100 million to build a minimum efficient scale production plant (or about \$10 million annually on an amortized basis). Such a plant could produce about 100 million pounds of cereal per year. What would be the average fixed costs of this plant if it ran at capacity? Each year, U.S. breakfast cereal makers sell about 3 billion pounds of cereal. What would be the average fixed costs if the cereal maker captured a 2 percent market share? What would be its cost disadvantage if it achieved only a 1 percent share? If prior to entering the market, the firm contemplates achieving only a 1 percent share, is it doomed to such a large cost disparity?
5. The European Union has banned virtually all tariffs for trade among member nations. How is this likely to affect specialization by firms located in EU countries?
6. Historically, product markets were dominated by large firms and service markets by small firms. This seems to have reversed itself somewhat in recent years. What factors might be at work?
7. Best Buy recently redesigned its stores so that customers waiting to make purchases must stand in a single queue and wait for the next available cashier, rather than queue up at

separate cashiers. How does this relate to inventory economies of scale? (*Hint:* What is inventoried?)

8. In the past few years, several American and European firms opened "hypermarts," enormous stores that sell groceries, household goods, hardware, and other products under one roof. What are the possible economies of scale that might be enjoyed by hypermarts? What are the potential diseconomies of scale?

9. Some state governors have proposed purchasing prescription drugs on behalf of state residents, on the grounds that by pooling purchasing power, they can obtain deep discounts. What advice would you give to governors to improve their chances of obtaining low prices?

10. Suppose you wanted to quantify a firm's learning experience. One possible measure is the firm's lifetime cumulative output. What are the advantages and disadvantages of this measure? Can you offer a superior alternative measure?

11. During the 1990s, firms in the Silicon Valley of northern California experienced high rates of turnover as top employees moved from one firm to another. What effect do you think this turnover had on learning by doing at individual firms? What effect do you think it had on learning by the industry as a whole?

Appendix

Using Regression Analysis to Estimate the Shapes of Cost Curves and Learning Curves

Estimating Cost Curves

Suppose that you had the following cost and output data for three chain saw manufacturing plants:

Plant	Annual Output	Average Cost
1	10,000	$50
2	20,000	$47
3	30,000	$45

Average costs apparently fall as output increases. It would be natural to conclude from this pattern that there are economies of scale in chain saw production. One might be tempted to recommend to the managers of plants 1 and 2 that they expand output (perhaps by building larger plants) so as to lower their average costs.

Just how confident should we be about the presence of scale economies in this instance? Put another way, what if the differences in costs at the three plants have nothing to do with scale economies? For example, plant 3 may be located in a region where labor costs are unusually low. If this is the case, then the other plants may have nothing to gain by expanding at their present locations. To be confident that the cost/output relationship truly reflects scale economies, alternative explanations need to be ruled out.

These are the ideas underlying *regression analysis of cost functions*. Regression analysis is a statistical technique for estimating how one or more factors affect some variable of interest. For cost functions, the variable of interest is average cost, and the factors may include output, wage rates, and other input prices.

To illustrate, suppose that we suspect that the average cost function is a quadratic function of the volume of output:

$$AC = \beta_0 + \beta_1 Q + \beta_2 Q^2 + \beta_3 w + noise$$

where Q denotes production volume (e.g., number of standard-sized chain saws produced per year), w denotes the local wage rate, and *noise* represents all of the other factors that affect the level of cost that cannot be measured and that are not explicitly included in the analysis.

We can interpret the cost function as follows: The average cost at any particular plant is equal to some function of plant output, plus a function of wage rates, plus noise. We expect β_3 to be positive, as higher wages contribute to higher costs. We expect β_1 to be negative, suggesting that as output rises, average costs fall. We expect β_2 to be small and positive. Thus, at large levels of output (and therefore at very large levels of output squared), average costs may start to level off or even increase, as the positive effect of $\beta_2 Q^2$ offsets or dominates the negative effect of $\beta_1 Q$. It is the combination of β_1, whose negative slope indicates economies of scale, and β_2, whose positive slope indicates diseconomies of scale, that produces the characteristic parabolic, or U, shape of the average cost function.

Finally, the noise term represents variation in costs due to factors other than size and wage rates. If we had good information about sources of that variation, we could directly include additional variables in the cost function. Otherwise, we are forced to accept that our cost function is necessarily imprecise. Regression analysis "fits" the cost function to actual cost/output data. In other words, regression provides estimates of the parameters β_1, β_2, and β_3 as well as the precision of these estimates.

There is a large literature on the estimation of cost functions. Cost functions have been estimated for various industries, including airlines, telecommunications, electric utilities, trucking, railroads, and hospitals.[19] Most of these studies estimate functional forms for the average cost function that are more complicated than the simple quadratic functions discussed here. Nevertheless, the basic ideas underlying these more sophisticated analyses are those described here, and these studies can be used to derive estimates of minimum efficient scale.

Estimating Learning Curves

Regression analysis may also be used to estimate learning curves. To do this, it is often convenient to estimate an equation with the following functional form:

$$\log AC = \alpha_0 + \alpha_E \log E + \alpha_1 \log X_1 + \cdots + \alpha_n \log X_n + noise$$

where "log" represents the natural logarithm, E denotes cumulative production volume, $X_1, \ldots, X_n$ denote the levels of cost drivers other than cumulative production volume that affect average cost (e.g., scale, capacity utilization, input prices, and so forth), and *noise* denotes the impact of factors that cannot be measured and are thus not included in the analysis. These other cost drivers are included in the equation to distinguish between cost reductions that are due to learning and cost reductions that are due to economies of scale or favorable positions on other cost drivers. The parameter α_E is the percentage change in average cost per 1 percent change in cumulative experience, and α_i is the percentage change in average cost per 1 percent change in cost driver X_i. Because logarithms are used in the preceding equation, the estimated coefficients are elasticities.

Endnotes

[1]If you do not understand why this must be so, consider this numerical example. Suppose that the total cost of producing five bicycles is $500. The *AC* is therefore $100. If the *MC* of the sixth bicycle is $70, then total cost for six bicycles is $570 and *AC* is $95. If the *MC* of the sixth bicycle is $130, then total cost is $630 and *AC* is $105. In this example (and as a general rule), when $MC < AC$, *AC* falls as production increases, and when $MC > AC$, *AC* rises as production increases.

[2]Prahalad, C. K., and G. Hamel, "The Core Competence of the Corporation," *Harvard Business Review*, May–June 1990, pp. 79–91; Stalk, G., P. Evans, and L. Shulman, "Competing on Capabilities: The New Rules of Corporate Strategy," *Harvard Business Review*, March–April 1992, pp. 57–69; Itami, H., *Mobilizing Indivisible Assets*, Cambridge, MA, Harvard University Press, 1987.

[3]The opportunity cost is the best return that the investor could obtain if he or she invested a comparable amount of money in some other similarly risky investment. In this example, we have assumed, for simplicity, that the production line never depreciates and thus lasts forever. See Chapter 1 for further discussion.

[4]Levine, A., "Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry," mimeo, Stanford University, 2008.

[5]Newhouse, J., et al., "Does the Geographic Distribution of Physicians Reflect Market Failure?" *Bell Journal of Economics*, 13(2), 1982, pp. 493–505.

[6]Baumgardner, J., "What Is a Specialist Anyway?" Mimeo, Duke University, 1991.

[7]Lynk, W., "The Creation of Economic Efficiency in Hospital Mergers," *Journal of Health Economics*, 14(6), 1995, pp. 507–530.

[8]The name *cube-square rule* comes from the fact that the volume of a cube is proportional to the cube of the length of its side, whereas the surface area is proportional to the square of that length.

[9]See Cockenboo, L., "Production Functions and Cost Functions: A Case Study," in Mansfield, E. (ed.), *Managerial Economics and Operations Research*, 5th ed., New York, Norton, 1987.

[10]DiMasi, J., et al., "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics*, 10(2), 1991, pp. 107–142.

[11]Henderson, R., and I. Cockburn, "Scale, Scope, and Spillovers: Determinants of Research Productivity in the Pharmaceutical Industry," *RAND Journal of Economics*, 27(1), 1996, pp. 32–59.

[12]Danzon, P., A. Epstein, and S. Nicholson, 2004, "Mergers and Acquisitions in the Pharmaceutical and Biotech Industries," NBER Working Paper 10536.

[13]Milgrom, P., and J. Roberts, "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review*, 80(6), 1990, pp. 511–528.

[14]Much of the information in this example was obtained from Hsieh, J., and A. W. Rice, "AOL Time Warner Inc.: Piecing Together an Integrated Media Platform," Deutsche Banc Alex Brown, *Equity Research*, June 2001, p. 2.

[15]See, for example, *Perspectives on Experience*, Boston, Boston Consulting Group, 1970, for estimates of progress ratios for over 20 industries. See Lieberman, M., "The Learning Curve and Pricing in the Chemical Processing Industries," *RAND Journal of Economics*, 15(2), 1984, pp. 213–228, for learning curve estimates for 37 chemical products.

[16]The product life-cycle model has its origins in the marketing literature. See, for example, Levitt, T., "Exploit the Product Life Cycle," *Harvard Business Review*, November–December 1965, pp. 81–94.

[17]Benkard, C. L., "Learning and Forgetting: The Dynamics of Aircraft Production," mimeo, New Haven, CT, Yale University, 1998.

[18]Ramanarayanan, S., 2008, "Does Practice Make Perfect? An Empirical Analysis of Learning-by-Doing in Cardiac Surgery," mimeo, UCLA.

[19]John Panzar's article, "Determinants of Firm and Industry Structure," in Schmalensee, R., and R. D. Willig (eds.), *Handbook of Industrial Organization*, Amsterdam, North Holland, 1989, pp. 3–59, reviews some of the work on estimation of cost functions and provides many references to the literature.