CHAPTER 8

Profit Maximization and Competitive Supply

A cost curve describes the minimum cost at which a firm can produce various amounts of output. With this knowledge, we can now turn to a fundamental problem faced by every firm: How much should be produced? In this chapter, we will see how a perfectly competitive firm chooses the level of output that maximizes its profit. We will also see how the output choice of individual firms leads to a supply curve for the entire industry.

Our discussion of production and cost in Chapters 6 and 7 applies to firms in all kinds of markets, but in this chapter we discuss only firms in perfectly competitive markets. In a perfectly competitive market all firms produce the identical product, and each firm is so small in relation to the industry that its production decisions have no effect on market price. New firms can easily enter the industry if they perceive a potential for profit, and existing firms can exit the industry if they start losing money.

We begin by explaining why it makes sense to assume that firms maximize profit. We provide a rule for choosing the profit-maximizing output for firms in all markets-competitive or otherwise. Then we show7 how a competitive firm chooses its output in the short and long run. We see how this output choice changes as the cost of production or the prices of inputs change. In this way, we show how to derive the firm's *supply curve.* We then aggregate the supply curves of individual firms to obtain the *industry* supply curve. In the short run, firms in an industry choose which level of output to produce to maximize profit. In the long run, firms not only make output choices, but also decide whether to be in a market at all. We will see that the prospect of high profits encourages firms to enter an industry, while losses encourage them to leave.

8.1 *Profit Maximization*

In this section we analyze profit maximization by firms. First, we ask whether firms do indeed seek to maximize profit. Then, we describe a rule that any firm-whether in a competitive market or not-can use to find its profit-maximizing output level. Next, we consider the special case of a firm in a competitive market. We distinguish the demand curve facing a competitive firm from the market demand curve, and then use this information to describe the competitive firm's profit-maximization rule.

Do Firms Maximize Profit?

The assumption of profit maximization is frequently used in microeconomics because it predicts business behavior reasonably accurately and avoids unnecessary analytical complications. But whether firms do maximize profit has been controversial.

For smaller firms managed by their owners, profit is likely to dominate almost all the firm's decisions. In larger firms, however, managers who make day-to-day decisions usually have little contact with the owners (i.e., the stockholders). As a result, the owners of the firm cannot monitor the managers' behavior on a regular basis. Managers then have some leeway in how they run the firm and can deviate from profit-maximizing behavior to some extent.

Managers may be more concerned with goals such as revenue maximization to achieve growth or the payment of dividends to satisfy shareholders than with profit maximization. Managers might also be overly concerned with the firm's short-run profit (perhaps to earn a promotion or a large bonus) at the expense of its longer-run profit, even though long-run profit maximization better serves the interests of the stockholders.[1] (We discuss the implications of differences between the incentives of managers and owners of firms in greater detail in Chapter 17.)

Even so, managers' freedom to pursue goals other than long-run profit maximization is limited. If they do pursue such goals, shareholders or the boards of directors can replace them, or the firm can be taken over by new management. In any case, firms that do not come close to maximizing profit are not likely -to survive. Firms that do survive in competitive industries make long-run profit maximization one of their highest priorities.

Thus, our working assumption of profit maximization is reasonable. Firms that have been in business for a long time are likely to care a lot about profit,

[1] To be more exact, maximizing the market value of the firm is a more appropriate goal than profit maximization because market value includes the stream of profits that the firm earns over time. It is the stream of profits that is of direct interest to the stockholders.

whatever else their managers may appear to be doing. For example, a firm that subsidizes public television may seem public-spirited and altruistic. Yet this beneficence is likely to be in the long-run financial interest of the firm because it generates goodwill for the firm and its products.

8.2 *Marginal Revenue, Marginal Cost, and Profit Maximization*

Let's begin by looking at the profit-maximizing output decision for *any* firm, whether the firm operates in a perfectly competitive market or is one that can influence price. Since profit is the difference between (total) revenue and (total) cost, to find the firm's profit-maximizing output level, we must analyze its revenue. Suppose that the firm's output is q, and that it obtains revenue R. This revenue is equal to the price of the product P times the number of units sold: $R = Pq$. The cost of production C also depends on the level of output. The firm's profit is the difference between revenue and cost:

$$\pi(q) = R(q) - C(q)$$

(Here we show explicitly that π, R, and C depend on output. Usually we will omit this reminder.)

To maximize profit, the firm selects the output for which the difference between revenue and cost is the greatest. This is shown in Figure 8.1. Revenue $R(q)$ is a curved line, which accounts for the possibility that an increased output may be accomplished only with a lower price. The slope of the line, which shows how much revenue increases when output increases by one unit, is *marginal revenue.* Because there are fixed and variable costs, C(cf) is not a straight line; its slope,which measures the additional cost associated with an additional unit of output,is *marginal cost.* $C(q)$ is positive when output is zero because there is a fixed cost in the short run.

For low levels of output, profit is negative-revenue is insufficient to cover fixed .and variable costs. (Profit is negative when $q = 0$ because of fixed cost.) Here marginal revenue is greater than marginal cost, which tells us that increases in output will increase profit. As output increases,profit eventually becomes positive (for q greater than q_0) and increases until output reaches q^* units. Here the marginal revenue and marginal cost are equal, and q^* is the profit-maximizing output. Note that the vertical distance between revenue and cost, AB, is greatest at this point; equivalently '$\pi(q)$) reaches its peak. Beyond q^* units of production, marginal revenue is less than marginal cost, and profit falls, reflecting the rapid increase in the total cost of production.

To see why q^* maximizes profit another way, suppose output is less than q^*. Then, if the firm increases output slightly, it will generate more revenues than costs. In other words, the marginal revenue (the additional revenue from pro-

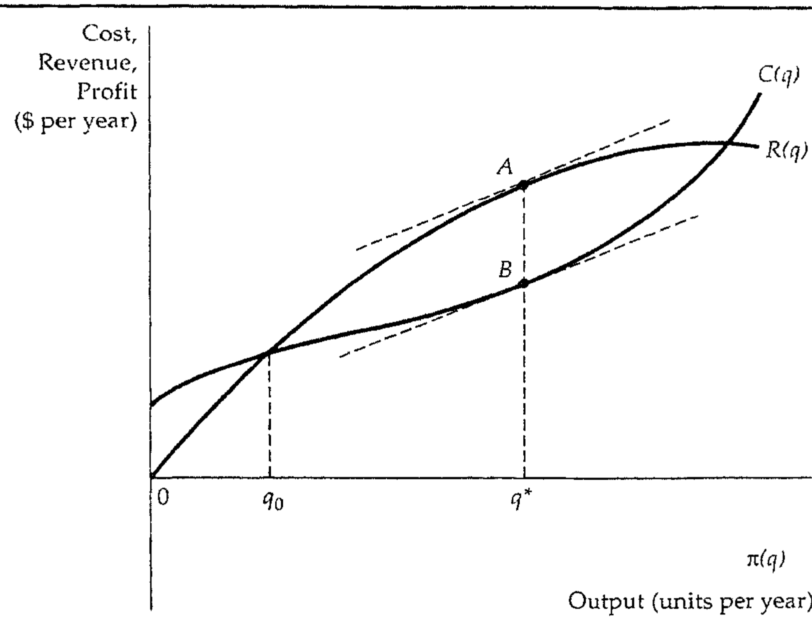

FIGURE 8.1 Profit Maximization in the Short Run. A firm chooses output q^*, so that profit, the difference AB between revenue R and cost C, is maximized. At that output marginal revenue (the slope of the revenue curve) is equal to marginal cost (the slope of the cost curve).

ducing one more unit of output) is greater than marginal cost. Similarly, when output is greater than q^*', marginal revenue is less than the marginal cost. Only when marginal revenue and marginal cost are equal has profit been maximized.

The rule that profit is maximized when marginal revenue is equal to marginal cost holds for all firms, whether competitive or not. This important rule can also be derived algebraically. Profit, $\pi = R - C$, is maximized at the point at which an additional increment to output just leaves profit unchanged (i.e., $\Delta\pi/\Delta q = 0$):

$$\Delta\pi/\Delta q = \Delta R/\Delta q - \Delta C/\Delta q = 0$$

$\Delta R/\Delta q$ is marginal revenue MR and $\Delta C/\Delta q$ is. marginal cost MC. Thus, we conclude that profit is maximized when MR - MC = 0, so that

$$\boxed{\text{MR}(q) = \text{MC}(q)}$$

Demand and Marginal Revenue for a Competitive Firm

Because each firm in a competitive industry sells only a small fraction of the entire industry sales, *how much output the firm decides to sell will have no effect on the market price of the product.* The market price is determined by the industry demand and supply curves. Therefore, the competitive firm is a *price taker:* It knows that its production decision will have no effect on the price of the product. For example, when a farmer is deciding how many acres of wheat to plant in a given year, he can take the market price of wheat as given. That price will not be affected by his acreage decision.

Often we will want to distinguish between market demand curves and the demand curves that individual firms face. In this chapter we will denote market output and demand by capital letters (Q and D), and the firm's output and demand by lower-case letters (q and d).

Because the firm is a price taker, the revenue curve R(q) is a straight line, since for a given P, revenue increases proportionately with output. In addition, *the demand curve d facing an individual competitive firm is given by a horizontal line.* In Figure 8.2a, the farmer's demand curve corresponds to a price of $4 per bushel of wheat. The horizontal axis measures the amount of wheat that the farmer can sell, and the vertical axis measures the price. Compare the demand curve facing the firm (in this case the farmer) in Figure 8.2a, with the market demand curve D in Figure 8.2b.

The market demand curve shows how much wheat *all consumers* will buy at each possible price. The market demand curve is downward sloping because consumers buy more wheat at a lower price. The demand curve facing the

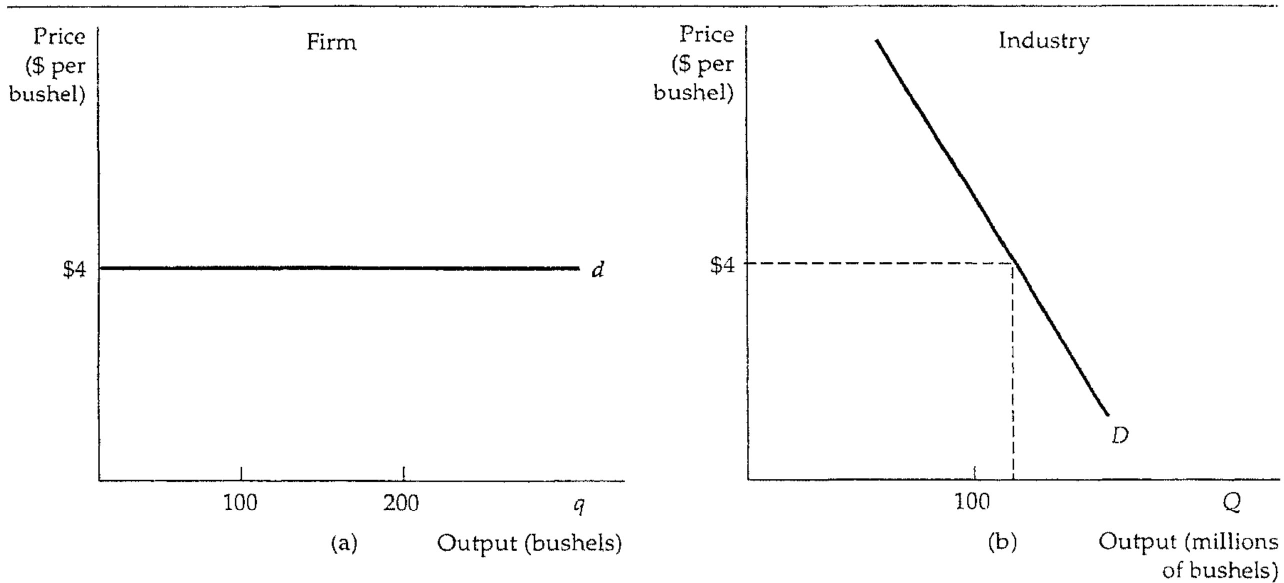

FIGURE 8.2 Demand Curve Faced by a Competitive Firm. A competitive firm supplies only a small portion of the total output of all the firms in an industry. Therefore, the firm takes the market price of the product as given,choosing its output on the assumption that the price will be unaffected by the output choice. In (a) the demand curve facing the firm is perfectly elastic, even though the market demand curve in (b) is downward sloping.

firm, however, ishorizontal because the firm's sales will have no effect on the price. Suppose the firm increased its sales from 100 to 200 bushels of wheat. This would have almost no effect on the market because the industry output of wheat is 100 million bushels at $4 per bushel. Price is determined by the interaction of all firms and consumers in the market, not by the output decision of a single firm.

When an individual firm faces a horizontal demand curve, it can sell an additional unit of output without lowering price. As a result, the total revenue increases by an amount equal to the price (one bushel of wheat sold for $4 yields additional revenue of $4). At the same time, the average revenue received by the firm is also $4 because each bushel of wheat produced will be sold at $4. Therefore, *the demand curve dfacing an individualfirm in a competitive market is both its average revenue curve and its marginal revenue curve. Along this demand curve, marginal revenue andprice are equal.*

Profit Maximization by a Competitive Firm

Because MR = P, and the demand curve facing a competitive firm is horizontal, the general profit-maximizing rule that applies to any firm can be simplified. A perfectly competitive firm should choose its output so that *marginal cost equals price:*

$$\boxed{MC(q) = MR = P}$$

Note that this is a rule for setting output, not price, since competitive firms take price as fixed. However, we show in Chapter 10 that the rule is a useful benchmark when we compare a noncompetitive firm's price with what the price would be if the market were competitive. The rule can help regulators decide what prices to set when they are regulating noncompetitive firms.

Because the choice of the profit-maximizing output by a competitive firm is so important, we will devote most of the rest of the chapter to analyzing it. We begin with the short-run output decision, and then move to the long run.

8.3 *Choosing Output in the Short Run*

How should the manager of a profit-maximizing firm choose a level of output over the short run, when the firm's plant size is fixed? In this section we show how afirmcan use information about revenue and cost to make a profit-maximizing output decision.

Short-Run Profit Maximization by a Competitive Firm

In the short run, a firm operates with a fixed amount of capital and must choose the levels of its variable inputs (labor and materials) to maximize profit. Figure 8.3 shows the firm's short-run decision. The average and marginal revenue

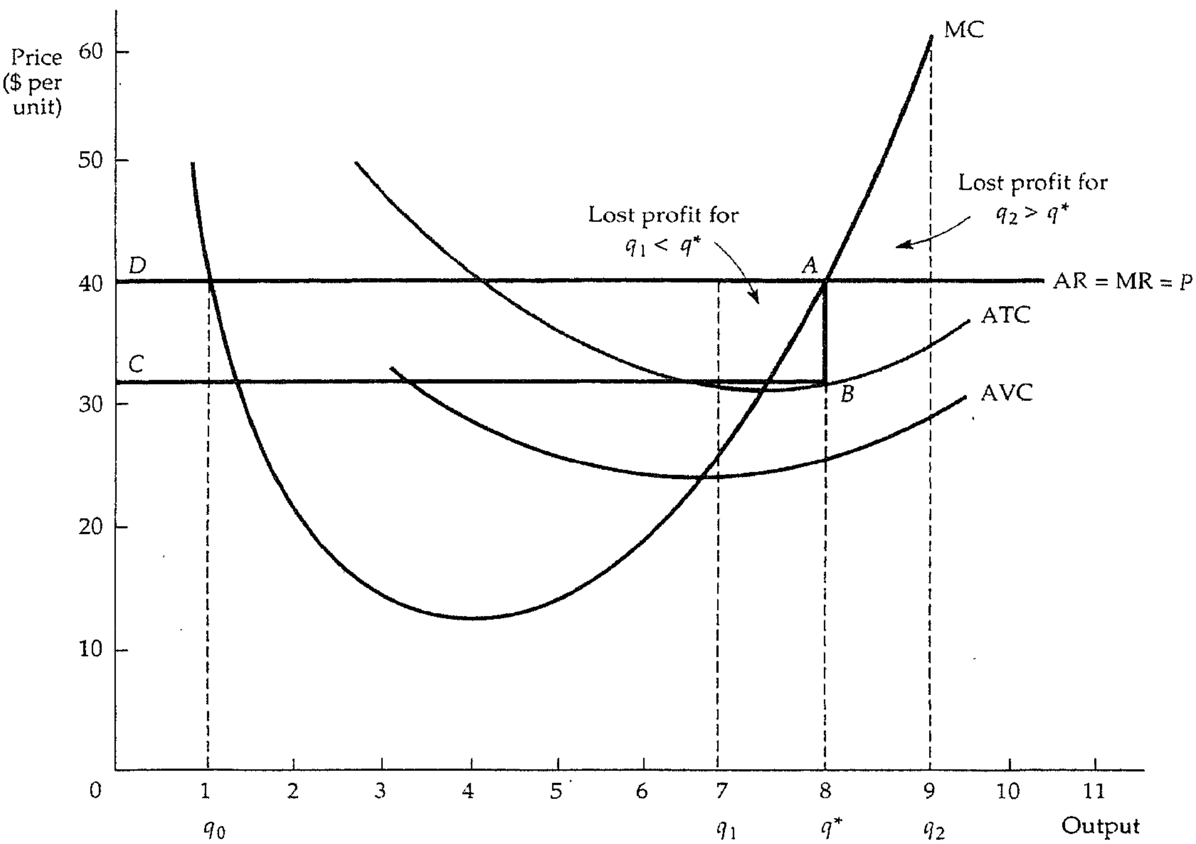

FIGURE 8.3 A Competitive Firm Making Positive Profit. In the short run, the competitive firm maximizes its profit by choosing an output q^* at which its marginal cost MC is equal *to* the price *P* (or marginal revenue MR) of its product. The profit of the firm is measured by the rectangle *ABCD*. Any lower output q_1 or higher output q_2 will lead to lower profit.

curves are drawn as horizontal lines at a price equal to $40. In this figure, we have drawn the average total cost curve ATC, the average variable cost curve AVC, and the marginal cost curve MC, so that we can see the firm's profit more easily.

Profit is maximized at point *A*, associated with an output $q^* = 8$ and a price of $40, because marginal revenue is equal to marginal cost at this point. At a lower output, say $q_1 = 7$, marginal revenue is greater than marginal cost, so profit could be increased by increasing-output. The shaded area between $q_1 = 7$ and q^* shows the lost profit associated with producing at q_1. At a higher output, say q_2 marginal cost is. greater than marginal revenue; thus, reducing output saves a cost that exceeds the reduction in revenue. The shaded area between q^* and $q_2 = 9$ shows the lost profit associated with producing at q_2.

The MR and MC curves cross at an output of q_0 as well as q^*. At q_0, however, profit is clearly not maximized. An increase in output beyond q_0 increases profit because marginal cost is well below marginal revenue. So the condition for profit maximization is that *marginal revenue equals marginal cost at a point at which the marginal cost curve is rising.*

The Short-Run Profitability of a Competitive Firm

Figure 8.3 also shows the competitive firm's short-run profit. The distance *AB* is the difference between price and average cost at the output level q^*, which is the average profit per unit of output. Segment *BC* measures the total number of units produced; Therefore, rectangle *ABCD* is the firm's total profit.

A firm need not always earn a profit in the short run, as Figure 8.4 shows. The major difference from Figure 8.3 is the higher fixed cost of production. This raises average total cost but does not change the average variable cost and

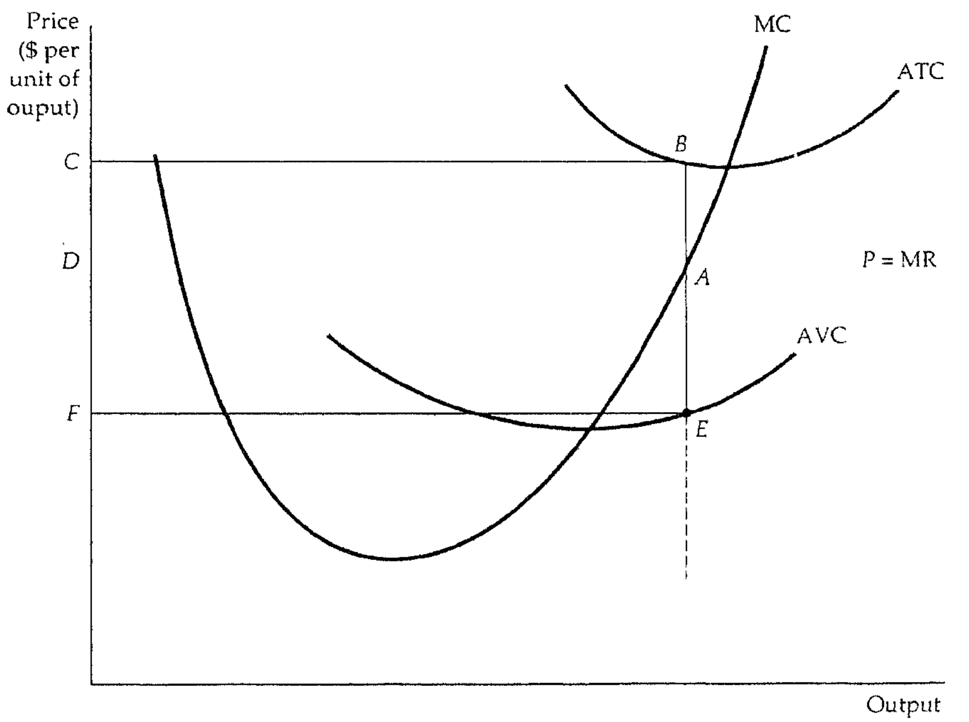

FIGURE 8.4 A Competitive Firm Incurring Losses. In the short run, a competitive firm may produce at a loss (because its fixed cost is high), if it can still generate revenues that more than cover its variable cost. The firm minimizes its losses by producing at q^*, with losses *ABCD*. If the firm were to shut down, it would incur even greater losses equal to the fixed cost of production *CBEF*.

marginal cost curves. At the profit-maximizing output q^*, the price P is less than average cost, so that line segment AB measures the average *loss from* production. Likewise, the shaded rectangle ABCD now measures the firm's total loss.

Why doesn't a firm that earns a loss leave the industry entirely? A firm might operate at a loss *in the short run* because it expects to earn a profit in the future as the price of its product increases or the cost of production falls. In fact, a firm has two choices in the short run: It can produce some output, or it can shut down its production temporarily. It will choose the more profitable (or the less unprofitable) of the two alternatives. In particular, *a firm will find it profitable to shut down (produce no output) when the price of its product is less than the minimum average variable cost.* In this situation, revenues from production will not cover variable costs, and losses will increase.

Figure 8.4 illustrates the case in which some production is appropriate. The output q^* is at the point where short-run losses are minimized. It is cheaper in this case to operate at q^* rather than to produce no output because price exceeds average variable cost at q^*. Each unit produced yields more revenue than cost, thereby generating higher profit *than* if the firm were to produce nothing. (Total profit is still negative, however, because the *fixed cost* is high.) Line segment AE measures the difference between price and average variable cost, and rectangle $AEFD$ measures the additional profit that can be earned by producing at q^* rather than at 0.

To see this another way, recall that the difference between average total cost ATC and average variable cost AVC is average fixed cost AFC. Therefore, in Figure 8.4, line segment B£ represents the average fixed cost, and rectangle *CBEF* represents the total fixed cost of production. When the firm produces no output, its loss is equal to its total fixed cost *CBEF*. But when it produces at q^*, its loss is reduced to the rectangle $ABCD$. Fixed cost, which is irrelevant to the firm's production decision in the short run, is crucial when determining whether the firm ought to leave the industry in the long run.

To summarize: The competitive firm produces no output if price is less than minimum average variable cost. When it does produce, it maximizes profit by choosing the output level at which price is equal to marginal cost. At this output level, profit is positive if price is greater than average total cost. The firm may operate at a loss in the short run. However, if it expects to continue to lose money over the long run, it will go out of business.

EXAMPLE 8.1 SOME COST CONSIDERATIONS FOR MANAGERS

The application of the rule that marginal revenue should equal marginal cost depends on the manager's ability to estimate marginal cost.[2] To obtain useful measures of cost, managers should keep three guidelines in mind.

[2]This example draws on the discussion of costs and managerial decision making in Thomas Nagle, *The Strategy and Tactics of Pricing* (Englewood Cliffs, N.J.: Prentice-Hall, 1987), Chapter 2.

First, when possible, *average variable cost should not be used as a substitute for marginal cost.* When marginal and average cost are nearly constant, there is little difference between them. However, when marginal and average cost are increasing sharply, the use of average variable cost can be misleading when deciding how much to produce. Suppose, for example, that a company has the following cost information:

Current output:	100 units per day, of which 25 units are produced during overtime
Materials cost:	\$500 per day
Labor cost:	\$2000 per day (regular) plus \$1000 per day (overtime)

Average variable cost is easily calculated-it is the labor and materials cost (\$3500) divided by 100 units per day, or \$35 per unit. But the appropriate cost is marginal cost, which could be calculated as follows: Materials cost per unit is likely to be constant whatever the output level, so that marginal materials cost is \$500/100 = \$5 per unit. Since the marginal cost of labor is likely to involve overtime work only, it is obtained by noting that 25 of the 100 units were produced during the overtime period. The average overtime pay per unit of production, \$1000/25 = \$40 per unit, provides a good estimate of the marginal cost of labor. Therefore, the marginal cost of producing an additional unit of output is \$45 per unit (the marginal materials cost plus the marginal labor cost); this is much larger than the average variable cost of \$35. If the manager relied on average variable cost, too much output would be produced.

Second, *a single item on a firm's accounting ledger may have two components, only one of which involves marginal cost.* Suppose, for example, that a manager is trying to cut back production. She reduces the number of hours that some employees work and lays off others. But the salary of an employee who is laid off may not be an accurate measure of the marginal cost of production when cuts are made because union contracts often require the firm to pay laid-off employees part of their salary. In this case, the marginal cost of increasing production is not the same as the savings in marginal cost when production is decreased. The savings is the labor cost after the required layoff salary has been subtracted.

Third, *all opportunity costs should be included in determining marginal cost.* Suppose a department store wants to sell children's furniture. Instead of building a new selling area, the manager decides to use part of the third floor, which had been used for appliances, for the furniture. The marginal cost of this space is the profit that would have been earned had the store continued to sell appliances there, per unit of furniture sold. This opportunity cost measure may be much greater than what the store actually paid for that part of the building.

These three guidelines can help a manager to measure marginal cost correctly. Failure to do so can cause production to be too high or too low, and thereby reduce profit.

8.4 *The Competitive Firm's Short-Run Supply Curve*

A *supply curve* for a firm tells us how much output it will produce at every possible price. We have seen that competitive firms will increase output to the point at which price is equal to marginal cost, but they will shut down if price is below average variable cost. Therefore, for positive output the firm's supply curve is the portion of the marginal cost curve that lies above the average variable cost curve. Since the marginal cost curve cuts the average variable cost curve at its minimum point (recall our discussion in Chapter 7 of marginal and average cost), *the firm's supply curve is its marginal cost curve above the point of minimum average variable cost.* For any P greater than minimum AVC, the profit-maximizing output can be read directly from the graph. At a price PI in Figure 8.5, for example, the quantity supplied will be q_1, and at P_2 it will be q_2. For' P less than (or equal to) minimum AVC, the profit-maximizing out-

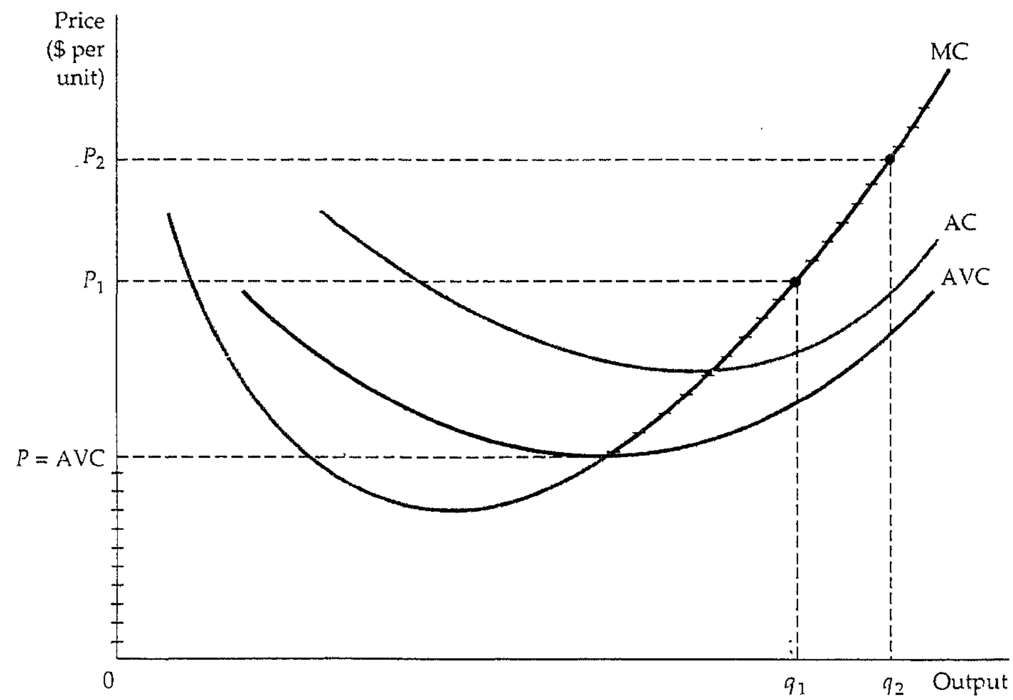

FIGURE 8.5 The Short-Run Supply Curve of a Competitive Firm. In the short run, the firm chooses its output so that marginal cost MC is equal to price, so long as it covers its variable cost of production. The short-run supply curve is given by the cross-hatched portion of the marginal cost curve.

put is equal to zero. In Figure 8.5 the entire supply curve is the cross-hatched portion of the vertical axis and the marginal cost curve.

Short-run supply curves for competitive firms slope upward for the same reason that marginal cost increases-the presence of diminishing returns to one or more factors of production. As a result, an increase in the market price will induce those firms already in the market to increase the quantities they produce. The higher price makes the additional production profitable and also increases the firm's *total* profit because it applies-to all units that the firm produces.

The Firm's Response to an Input Price Change

When the price of its product changes, the firm changes its output level, so that the marginal cost of production remains equal to the price. Often, however, the product price changes at the same time that the prices of *inputs* change. In this section we show how the firm's output decision changes in response to a change in the prices of one of the firm's inputs.

Figure 8.6 shows a firm's marginal cost curve that is initially given by MC_1 when the firm faces a price of $5 for its product. The firm maximizes its profit by producing an output of q_1. Now suppose the price of one of the firm's inputs increases. This causes the marginal cost curve to shift upward from MQ to MQ because it now costs more to produce each unit of output. The new profit-maximizing output is q_2 at which $P = MC_2$. Thus, the higher input price causes the firm to reduce its output.

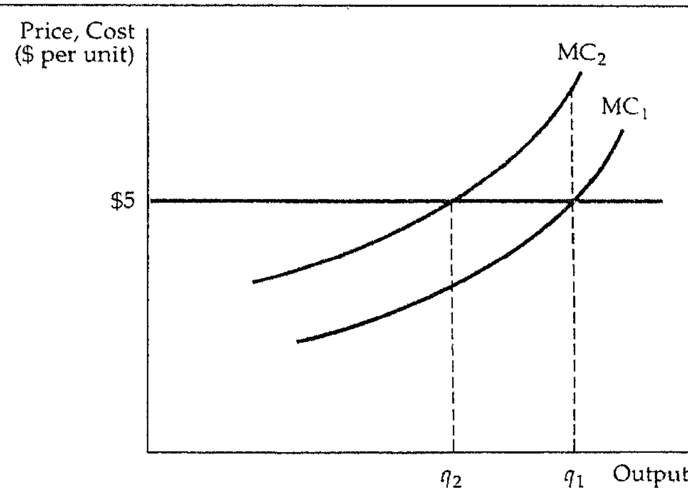

FIGURE 8.6 The Response of a Firm to a Change in Input Price. When the marginal cost of production for a firm increases (from MCi to MC2), the level of output that maximizes profit falls (from q_1 to q_2).

If the firm had continued to produce q_1, *it* would have incurred a loss on the last unit of production. In fact, all production beyond q_2 reduces profit. The shaded area in the figure gives the total savings to the firm (or equivalently, the reduction in lost profit) associated with the reduction in output from q_1 to q_2

EXAMPLE 8.2 THE SHORT-RUN PRODUCTION OF PETROLEUM PRODUCTS

Suppose you are managing an oil refinery and you decide to produce a particular combination of refinery products, including gasoline, jet fuel, and residual fuel oil for home heating. A lot of crude oil is available, but the amount that you refine depends on the capacity of the refinery and the cost of production. How much of the product mix should you produce each day?[3]

Information about the refinery's marginal cost of production is essential for this decision. Figure 8.7 shows the short-run marginal cost curve (SMC). Marginal cost increases with output, but in a series of uneven segments rather than as a smooth curve. The increase is in segments because the refinery uses different processing units to turn crude oil into finished products. When a particular processing unit reaches capacity, output can be increased only by substituting a more expensive process. For example,gasoline can be produced from light crude oils rather inexpensively in a processing unit called a "thermal cracker," When this unit becomes full, additional gasoline can still be produced (from heavy as well as light crude oil) but at a higher cost. In Figure 8.7 the first capacity constraint comes into effect when production reaches about 9700 barrels a day. A second capacity constraint becomes important when production increases beyond 10,700 barrels a day.

Deciding how much output to produce now becomes relatively easy. Suppose the mix of refined products can be sold for $23 per barrel. Since the marginal cost of production is close to $24 for the first unit of output, at a price of $23 no crude oil should be run through the refinery. If, however, the price of the product mix is between $24 and $25, you should produce 9700 barrels a day (filling the thermal cracker). Finally, if the price is above $25, you should use the more expensive refining unit and expand production toward 10,700 barrels a day.

Because the cost function rises in steps, you know that your production decisions need not change much in response to small changes in the price of the product. You will typically utilize sufficient crude oil to fill the appropriate processing unit until price increases (or decreases) substantially. Then you need simply calculate whether the increased price warrants using an additional, more expensive processing unit.

[3] This example is based on James M. Griffin, "The Process Analysis Alternative to Statistical Cost Functions: An Application to Petroleum Refining," *American Economic Review 62.* (1972): 46-56. The numbers have been updated and applied to a particular refinery.

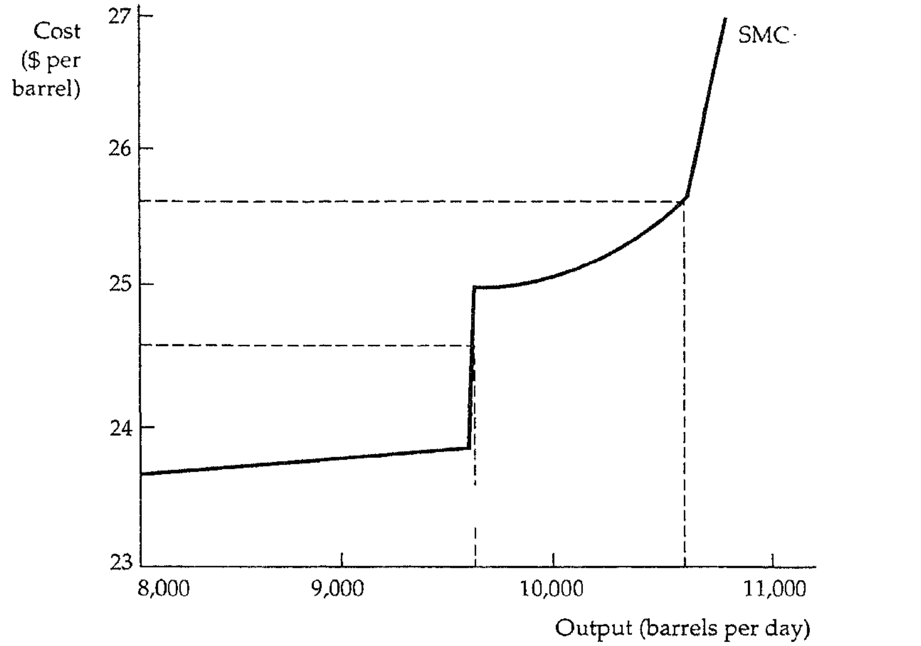

FIGURE 8.7 The Short-Run Production of Petroleum Products. The marginal cost of producing a mix of petroleum products from crude oil increases sharply at several levels of output as the refinery shifts from one processing unit to another. As a result, the output level can be insensitive to some changes in price and very sensitive to others.

8.5 *The Short-Run Market Supplu Curue*

The *short-run market supply curve shows* the amount of output that the industry will produce in the short run for every possible price. The industry's output is the sum of the quantities supplied by all the individual firms. Therefore, the market supply curve can be obtained by adding their supply curves. Figure 8.8 shows how this is done when there are only three firms, all of which have different short-run production costs. Each firm's marginal cost curve is drawn only for the portion that lies above its average variable cost curve. (We have shown only three firms to keep the graph simple, but the same analysis applies when there are many firms.)

At any price below P_1, the industry will produce no output because P_1 is the minimum average variable cost of the lowest-cost firm. Between P_1 and P_2, only firm 3 will produce, so the industry supply curve will be identical to that

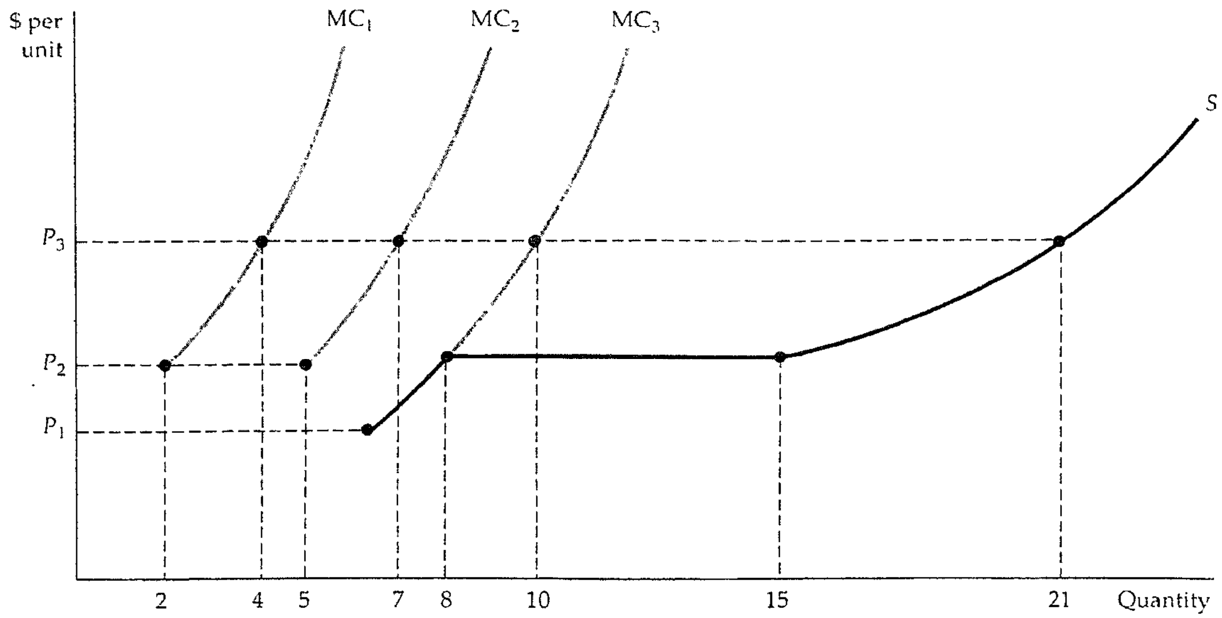

FIGURE 8.8 Industry Supply in the Short Run. The short-run industry supply curve is the horizontal summation of the supply curves of the individual firms. Because the third firm has a lower average variable cost curve than the first two firms, the market supply curve S begins at price P_1 and follows the marginal cost curve of the third firm MC3 until price equals P_2, where there is a kink. For all prices above P_2, the industry quantity supplied is the sum of the quantities supplied by each of the three firms.

portion of firm 3's marginal cost curve MC_3. At price P_2 the industry supply will be the. sum of the quantity supplied by all three firms. Firm 1 supplies 2 units, firm 2 supplies 5 units, and firm 3 supplies 8 units; thus, industry supply is 15 units. At price P_3 firm 1 supplies 4 units, firm 2 supplies 7 units, and firm 3 supplies 10 units; in total the industry supplies 21 units. Note that the industry supply curve is upward sloping but has a kink at price P_2. With many firms in the market, however, the kink becomes unimportant,so we usually draw industry supply as a smooth, upward-sloping curve.

Finding the industry supply curve is not always as simple as adding up a set of firm supply curves. As price rises, all firms in the industry expand their output. This additional output increases the demand for inputs to production and may lead to higher input prices. As we saw in Figure 8.6, increasing input prices shifts the firms' marginal cost curves upward. For example, an increased demand for beef could also increase demand for corn and soybeans (which are used to feed cattle), and thereby cause the prices of these crops to rise. In turn, the higher input prices would cause beef firms' marginal cost curves to shift upward: This lowers each firm's output choice (for any given market price) and causes the industry supply curve to be less responsive to changes in output price than it would otherwise be.

Elasticity of Market Supply

The price elasticity of market supply measures the sensitivity of industry output to market price. Recall from Chapter 2 that the elasticity of supply E_S is the percentage change in quantity supplied Q in response to a 1 percent change in price P:

$$E_S = (\Delta Q/Q)/(\Delta P/P)$$

Because marginal cost curves are upward sloping, the short-run elasticity of supply is always positive. When marginal costs increase rapidly in response to increases in output, the elasticity of supply is low. Firms are then capacity-constrained and find it costly to increase output. But, when marginal costs increase slowly in response to increases in output, supply is relatively elastic, and a small price increase induces firms to produce much more.

At one extreme is the case of *perfectly inelastic supply,* which arises when the industry's plant and equipment are so fully utilized that new plants must be built (as they will be in the long run) to achieve greater output. At the other extreme is the case of *perfectly elastic supply,* which arises when marginal costs are constant.

EXAMPLE 8.3 THE SHORT-RUN WORLD SUPPLY OF COPPER

In the short run, the shape of the market supply curve .for a mineral such as copper depends on how the cost of mining varies within and among/the world's major producers. Costs of mining, smelting, and refinmg copper differ because of differences in labor and transportation costs and differences in the copper content of the ore. Table 8.1 summarizes some of the relevant cost and production data for the largest copper-producing nations.[4]

These data can be used to plot the world supply curve for copper. The supply curve is a short-run curve because it takes the existing mines as fixed. Figure 8.9 shows how this curve is constructed for the six countries listed in the table. The complete world supply curve would, of course, incorporate data for all copper-producing countries. Also, note that the curve in Figure 8.9 is an approximation. The marginal cost number for each country is an average for all copper producers in that country. In the United States, for example, some producers had a marginal cost greater than 84 cents, and some less than 84 cents.

[4] The former Soviet Union is excluded because of data limitations. The source is the U.S. Department of the Interior, Bureau of Mines, *Minerals Yearbook,* 1985, Tables 4 and 31.

TABLE 8.1 The World Copper Industry (1990)

Country	Annual Production (thousand metric tons)	Marginal Cost (dollars per pound)
Canada	760	1.10
Chile	1550	.72
Peru	373	.98
United States	1550	.84
Zaire	390	.61
Zambia	430	.67

The lowest-cost copper is mined in Zaire, where the marginal cost of refined copper was about 61 cents per pound.[5] Curve MCz describes this marginal cost curve. The curve is horizontal until Zaire's capacity to mine copper is reached. Curve MCzm describes Zambia's supply curve (marginal cost is 67 cents per

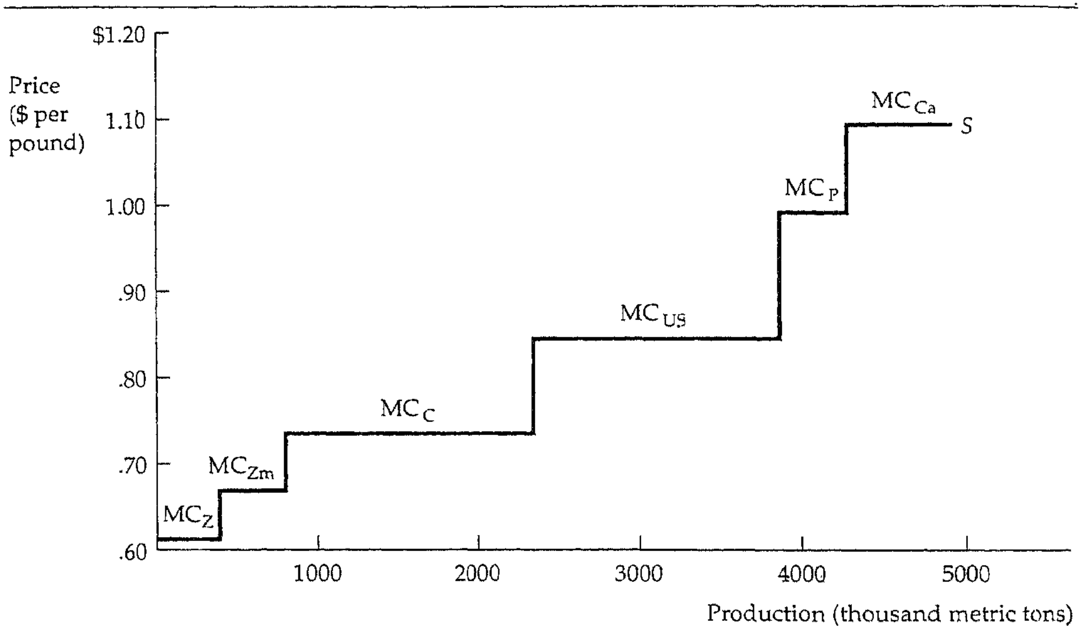

FIGURE 8.9 The Short-Run World Supply of Copper. The supply curve for world copper is obtained by summing the marginal cost curves for each of the major copper-producing countries. The supply curve slopes upward because the marginal cost of production ranges from a low of 61 cents per pound in Zaire to a high of $1.10 per pound in Canada.

[5] We are presuming that marginal and average costs of production are approximately the same.

pound). Likewise, curves MCc, MCus, MCp, and MCca represent the marginal cost curves for Chile, the United States, Peru, and Canada, respectively.

The world supply curve, denoted S, is obtained by summing each nation's supply curve horizontally. The slope and the elasticity of the supply curve depend on the price of copper. At relatively low prices, such as 70 to 85 cents per pound, the supply curve is quite elastic because small price increases lead to substantial increases in refined copper. But for higher prices, say above $1.10 per pound, the supply curve becomes quite inelastic because at such prices all producers would be operating at capacity.

Producer Surplus in the Short Run

In Chapter 4 we measured consumer surplus as the difference between the maximum that a person would pay for an item and its market price. An analogous concept applies to firms. If marginal cost is rising, the price of the product is greater than marginal cost for every unit produced except the last one. As a result, firms earn a surplus on all but the last unit of output. The *producer surplus* of a firm is the sum over all units produced of the difference between the market price of the good and the marginal cost of production. Just as consumer surplus measures the area below an individual's demand curve and above the market price of the product, producer surplus measures the area above a producer's supply curve and below the market price.

Figure 8.10 illustrates short-run producer surplus for a firm. The profit-maximizing output is q^*, where $P = \text{MC}$. Producer surplus is given by the shaded area under the firm's horizontal demand curve and above its marginal cost curve, from zero output to the profit-maximizing output q^*.

When we add the marginal costs of producing each level of output from 0 to q^*, we find that the sum is the total variable cost of producing q^*. Marginal cost reflects increments to cost associated with increases in output; since fixed cost does not vary with output, the sum of all marginal costs must equal the sum of the firm's variable costs. Thus, producer surplus can alternatively be defined as the difference between the firm's revenue and its total variable cost. In Figure 8.10 producer surplus is also therefore given by the rectangle *ABCD*, which equals revenue (0ABq*) minus Variable cost (0DCq^*).

Producer surplus is closely related to profit, but is not equal to it. Producer surplus is equal to revenue net of variable cost, while profit is equal to revenue net of *all* costs, both variable and fixed:

$$\text{Producer Surplus} = \text{PS} = R - \text{VC}$$

$$\text{Profit} = \pi = R - \text{VC} - \text{FC}$$

It follows that in the short run when the fixed cost is positive, producer surplus is greater than profit.

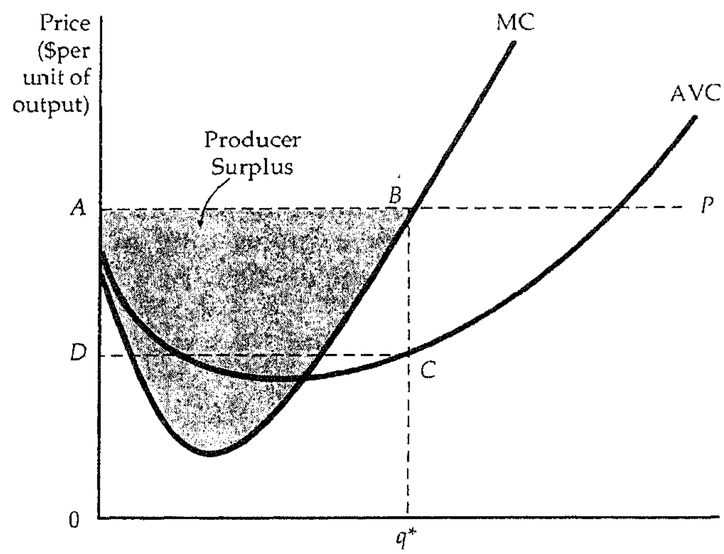

FIGURE 8.10 Producer Surplus for a Firm. The producer surplus for a firm is measured by the shaded area below the market price and above the marginal cost curve, between outputs 0 and q^* the profit-maximizing output. Alternatively, it is equal to rectangle *ABCD* because the sum of alt marginal costs up to q^* is equal to the variable costs of producing q^*.

The extent to which firms enjoy producer surplus depends on their costs of production. Higher-cost firms have less producer surplus, and lower-cost firms have more. By adding up all of the individual firms' producer surpluses, we can determine the producer surplus for a market. This can be seen in Figure 8.11. In the figure the market supply curve begins at the vertical axis at a point that represents the average variable cost of the lowest-cost firm in the market. Producer surplus is the area that lies below the market price of the product and above the supply curve between the output levels 0 and Q^*.

8.6 *Choosing Output in the Long Run*

In the long run, a firm can alter all its inputs, including the size of the plant. It can decide to shut down (i.e., to *exit* the industry) or to begin to produce a product for the first time (i.e., to *enter* an industry). Because we are concerned here with competitive markets, we allow *for free entry* and *free exit.* In other words, we are assuming that firms may enter or exit without any legal restriction or any special costs associated with entry.[6]

[6] In Chapter 10 we discuss examples of barriers to entry in an industry.

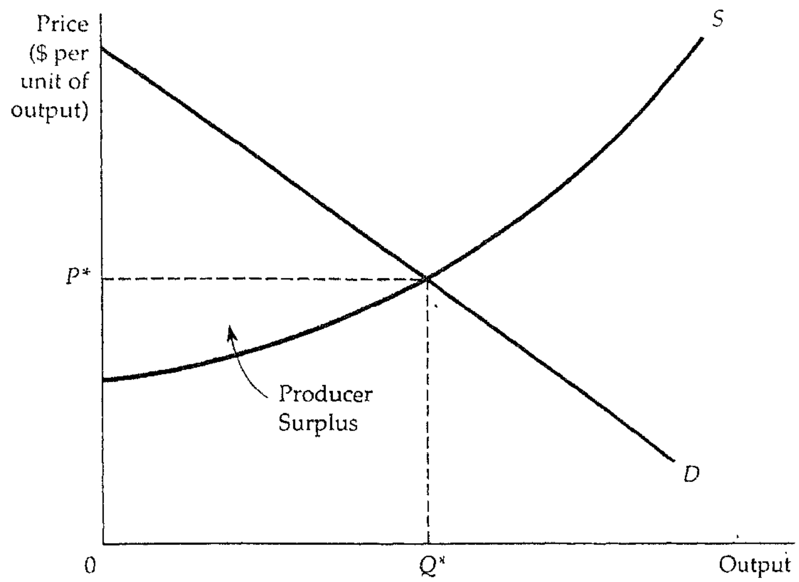

FIGURE 8.11 Producer Surplus for a Market. The producer surplus for a market is the area below the market price and above the market supply curve, between 0 and output Q^*.

Figure 8.12 shows how a competitive firm makes its long-run, profit-maximizing output decision. As in the short run, it faces a horizontal demand curve. (In Figure 8.12 the firm takes the market price of $40 as given.) Its short-run average (total) cost curve SAC and short-run marginal cost curve SMC are low enough for the firm to make a positive profit, given by rectangle *ABCD*, by producing an output of q_1 where SMC = P = MR. The long-run average cost curve LAC reflects the presence of economies of scale up to output level *qz* and diseconomies of scale at higher output levels. The long-run marginal cost curve LMC cuts the long-run average cost from below at q_2 the point of minimum long-run average cost.

If the firm believes the market price will remain at $40, it will want to increase the size of its plant to produce an output q3 at which its *long-run* marginal cost is equal to the $40 price. When this expansion is complete, the firm's profit margin will increase from *AB* to *EF*, and its total profit will increase from *ABCD* to *EFGD*. Output q_3 is profit-maximizing for the firm because at any lower output, say q_2 the-marginal revenue from additional production is greater than the marginal cost, so expansion is desirable.But at any output greater than q_3 marginal cost is greater than marginal revenue, so additional production would reduce profit. In summary, *the long-run output of a profit-maximizing competitive firm is where long-run marginal cost is equal to price.*

Note that the higher the market price, the higher the profit that the firm can earn. Correspondingly, as the price of the product falls from $40 to $30, so does the profit of the firm. At a price of $30, the firm's profit-maximizing

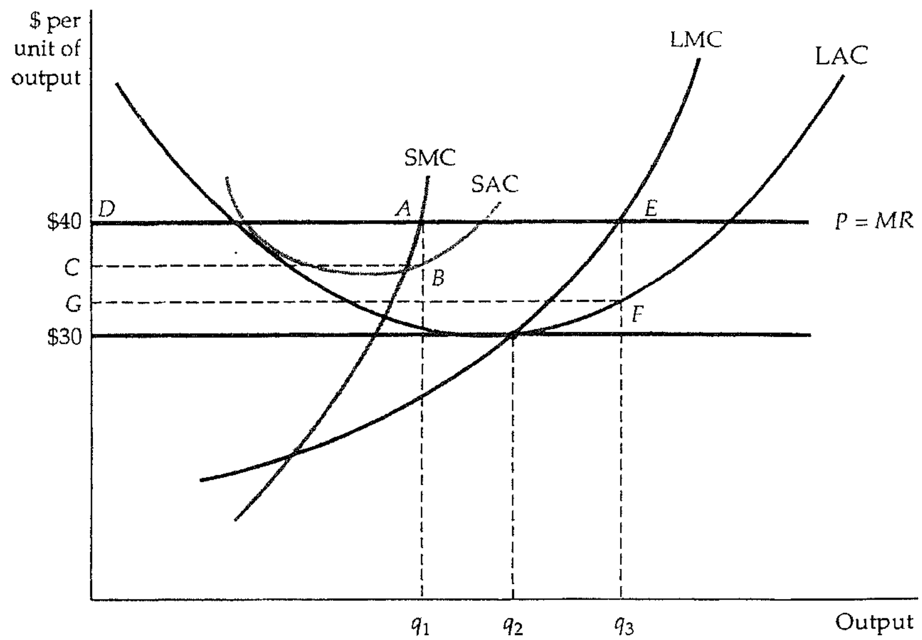

FIGURE 8.12 Output Choice in the Long Run. The firm maximizes its profit by choosing the output at which price is equal to long-run marginal cost LMC. In the diagram, the firm increases its profit from *ABCD* to *EFGD* by increasing its output in the long run.

output is q_2, the point of long-run minimum average cost. In this case, since P = ATC, the firm earns zero economic profit. As we show below, this means that investors in the firm earn a competitive return on their investment.

Zero Profit

As we saw in Chapter 7, it is important to distinguish between accounting profit and economic profit Accounting profit is measured by the difference. between the firm's revenues and costs, including actual outlays and depreciation expenses. Economic profit takes account of opportunity costs. One such opportunity cost is the return that the owners of the firm could make if their capital were used elsewhere. Suppose, for example, that the firm uses labor and capital inputs; its capital equipment has been purchased (and has been depreciated). The firm's accounting profit will equal its revenues R minus its labor costs wL, which is positive. However, its economic profit π equals its revenues R minus its labor costs wL and minus its opportunity cost of capital rK, measured by what it could rent the capital for in the market. Thus,

$$\pi = R - wL - rK = 0$$

A firm earning a negative economic profit should consider going out of business if it does not expect to improve its financial picture. However, a firm that earns zero economic profit need not go out of business because zero profit means the firm is earning a reasonable return on its investment. Of course, investors would like to earn a positive economic profi-that is what encourages entrepreneurs to develop and commercialize new ideas. But in competitive markets, as we will see, economic profits become zero in the long run. Zero economic profits signify not that the firms in the industry are performing poorly, but rather, that the industry is competitive.

Long-Run Competitive Equilibrium

Figure 8.12 shows how a $40 price induces a firm to increase its output and gives the firm a positive profit. Because profit is calculated net of the opportunity cost of investment, a positive profit means an unusually high return on investment. This high return causes investors to direct resources away from other industries and into this one-there will be *entry* into the market. Eventually the increased production associated with new entry causes the market supply curve to shift to the right, so that market output increases and the market price of the product falls.[7] Figure 8.13 illustrates this. In part (b) of the figure, the supply curve has shifted from S_1 to S_2, causing the price to fall from P_1 ($40) to P_2 ($30). In part (a), which applies to a single firm, the long-run average cost curve is tangent to the horizontal price line at output q_2.

When a firm earns zero economic profit, it has no incentive to exit the industry, and other firms have no special incentive to enter. A *long-run competitive equilibrium* occurs when three conditions hold. First, all firms in the industry are maximizing profit. Second, no firm has an incentive either to enter or exit the industry because all firms in the industry are earning zero economic profit. Third, the price of the product is such that the quantity supplied by the industry is equal to the quantity demanded by consumers.

The dynamic process that leads to long-run equilibrium creates a puzzle. Firms enter the market because of the opportunity to earn positive profit, and they exit because of losses. Yet, in long-run equilibrium, firms earn zero economic profit. Why do firms exit or enter if they know that eventually they will be no better or worse off than if they do nothing? The answer is that it can take a long time to reach a long-run equilibrium, and a substantial profit (or loss) can be made in the short run. The first firm to enter a profitable industry can earn much more short-run profit for its investors than can firms that enter later. Similarly, the first firm to exit an unprofitable industry can save its investors lots of money. Thus, the concept of long-run equilibrium tells us the *direction* that firms' behavior is likely to take. The idea of an eventual zero-profit, long-run equilibrium should not discourage a manager whose reward depends on the short-run profit that the firm earns.

[7] We discuss why the long-run supply curve might be upward sloping in the next section.

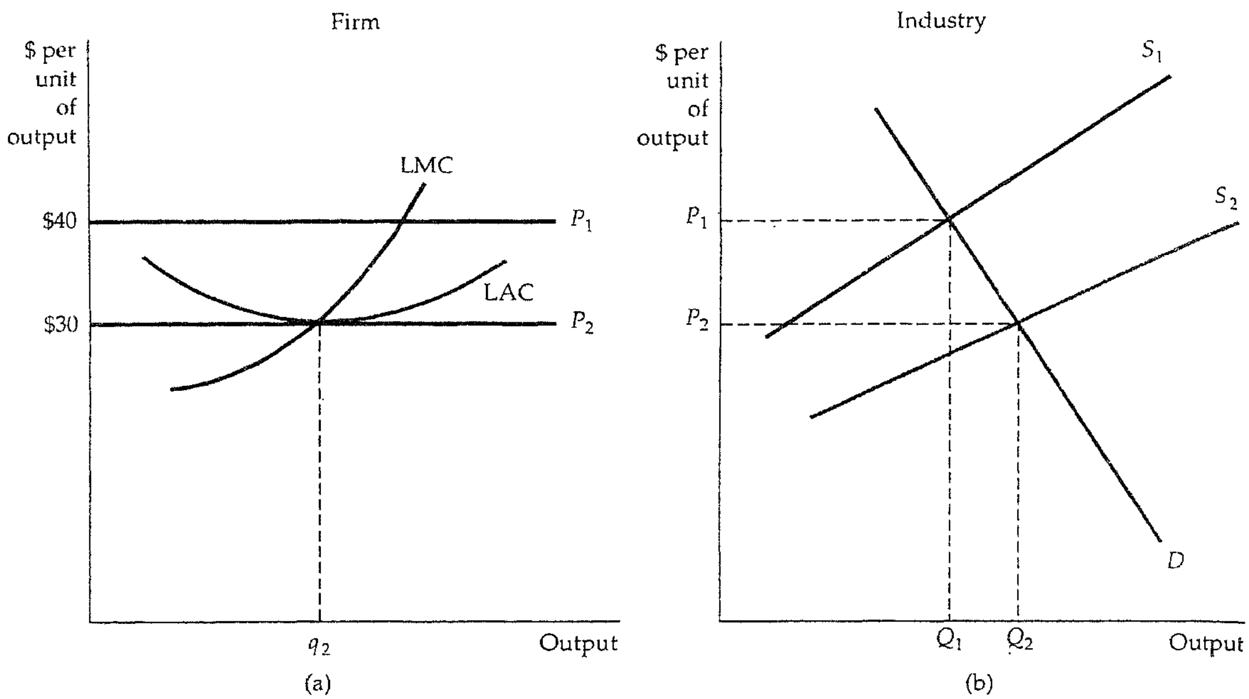

FIGURE 8.13 Long-Run Competitive Equilibrium. Initially the long-run equilibrium price of a product is $40 per unit, as shown in (b) as the intersection of demand curve D and supply curve S_1. In (a) it is shown that firms earn a positive profit because their long-run average cost reaches a minimum of $30 (at q_2). This positive profit encourages entry of new firms and causes a shift to the right in the supply curve to S_2. The long-run equilibrium occurs at a price of $30 because each firm earns zero profit, and there is no incentive to enter or exit the industry.

To see why all the conditions for long-run equilibrium must hold, assume that all firms have identical costs, and consider what happens if too many firms enter the industry in response to an opportunity for profit. Then the supply curve in Figure 8.13b will Shift further to the right, and price will fall below $30, say to $25. At that price, however, firms will lose money. As a result, some firms will exit the industry. Firms will continue to exit until the market supply curve shifts back to S_2. Only when there is no incentive to exit or enter the industry can a market be in long-run equilibrium.

Now suppose that all firms in the industry do not have identical cost curves. One firm has a patent or new idea that lets it produce at a lower average cost than all other firms. Then, it is consistent with long-run equilibrium for that firm to be earning a positive *accounting* profit (and to enjoy a higher producer surplus than other firms). As long as other investors and firms cannot acquire the patent or idea that lowers costs, they have no incentive to enter the industry. And as long as the process is particular to this product and this industry, the fortunate firm has no incentive to exit the industry. The distinction between accounting profit and economic profit is important here. If the new idea or invention is profitable,other firms in the industry will pay to use it. (Or they might attempt to buy the entire firm to acquire the patent.) The increased value of the patent thus represents an opportunity cost to the firm-it could sell the rights to the patent rather than use it. If all firms are equally efficient otherwise, once this opportunity cost is accounted for, the *economic* profit of the firm falls to zero.[8]

There are other instances in which firms earning positive accounting profit may be earning zero economic profit. Suppose, for example, that a clothing store happens to be located near a large shopping center. The additional flow of customers may substantially increase the store's accounting profit because the cost of the land is based on its historical cost. However, as far as economic profit is concerned, the cost of the land should reflect its opportunity cost, which in this case is its current market value. When the opportunity cost of land is included, the profitability of the clothing store is no higher than that of its competitors.

Thus, the condition that economic profit be zero is essential for the market to be in a long-run equilibrium. Positive economic profit, by definition, represents an opportunity for investors and an incentive to enter the industry. Positive accounting profit, however, may signal that firms already in the industry possess valuable assets, skills, or ideas, and this will not necessarily encourage entry by other firms.

Economic Rent

We have seen that some firms earn higher accounting profit than others because they have access to factors of production that are in limited supply; these might include land and natural resources, entrepreneurial skill, or other creative talent. What makes economic profit zero in the long run in these situations is the willingness of other firms to use the factors of production that are in limited supply. The positive accounting profits are therefore translated into *economic rent* that is earned by the scarce factors. *Economic rent* is defined as the difference between what firms are willing to pay for an input to production less the minimum amount necessary to buy that input. In-competitive markets, in both the short and the long run, economic rent is often positive, even though profit is zero.

For example, suppose that two firms in an industry own their land outright; the minimum cost of obtaining the land is zero. One firm is located on a river and can ship its products for $10,000 a year less than the other firm, which is

[8] If the firm with the patent is more efficient than other firms, then it will be earning a positive profit. But if the patent holder is less efficient, it should sell off the patent and go out of business.

inland. Then, the $10,000 higher profit of the first firm is due to the $10,000 per year economic rent associated with its river location. The rent is created because the land along the river is valuable, and otherfirms would be willing to pay for it. Eventually, the competition for this specialized factor of production will increase its value to $10,000. Land rent-the difference between $10,000 and the zero cost of obtaining the land-is also $10,000. Note that while the economic rent has increased, the economic profit of the firm on the river has become zero.

The zero economic profit tells the firm on the river that it should remain in the industry only if it is at least as efficient in production as other firms. It also tells possible entrants to the industry that entry will be profitable only if they can produce more efficiently than firms already producing.

Producer Surplus in the Long Run

When a firm is earning a positive accounting profit, but there is no incentive for other firms to enter or exit the industry, this profit must reflect economic rent. How then does rent relate to producer surplus? Recall that producer surplus measures the difference between the market price a producer receives and the marginal cost of production. Thus, in the long run, in a competitive market, *the producer surplus that a firm earns consists of the economic rent that it enjoys from all its scarce inputs*[9]

Suppose, for example, that a baseball team has a franchise that makes it the only team in a particular city. The team will earn a substantial accounting profit. This profit will include some economic rent because the team is more valuable with the franchise than it would be if entry into the local baseball market were unrestricted. The producer surplus earned by the baseball team would include its economic profit and the rent that reflects the difference between the current value of the team and what its value would be if an unlimited number of franchises were available.

Figure 8.14 shows that firms that earn economic rent earn the same economic profit as firms that do not earn rent. Part (a) shows the economic profit of a baseball team located in a city with several competing teams. The average price of a ticket is $7, and costs are such that the team earns zero economic profit. Part (b) shows the profit of a team with the same costs, but in a city with no competing teams. Because it is the only team in town, it can sell tickets for $10 apiece, and thereby earn an accounting profit of close to $3 on each ticket. However, the rent associated with the desirable location represents a. cost to the firm-an opportunity cost-because it could sell its franchise to another team. As a result, the economic profit in the city without competition is also zero.

[9] In a noncompetitive market producer surplus will reflect economic profit as well as economic rent.

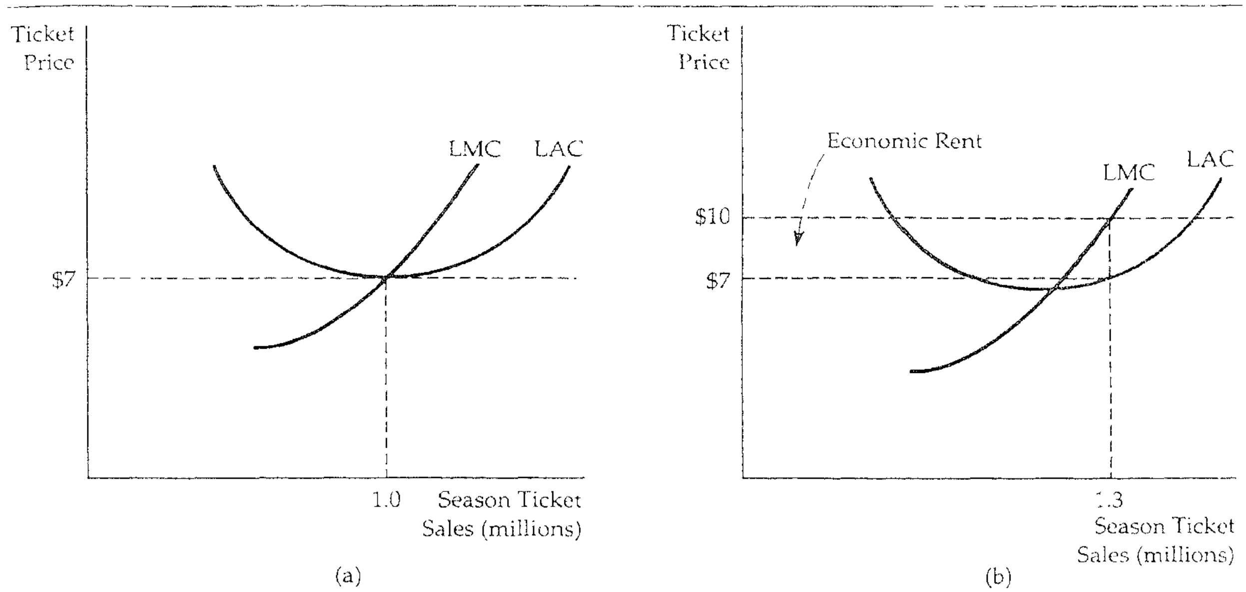

FIGURE 8.14 Firms Earn Zero .Profit in Long-Run Equilibrium. In long-run equilibrium, all firms earn zero economic profit. In (a) a baseball team in a city with other competitive sports teams sells enough tickets so that price ($7) is equal to marginal and average cost. In (b) there are no other competitors, so a $10 price can be charged. The team increases its sales to the point at which the average cost of production plus the average economic rent is equal to the ticket price. When the opportunity cost associated with owning the franchise is taken into account, the team earns zero economic profit.

8.7 *The Industry's Long-Run Supply Curve*

In our analysis of short-run supply, we first derived the firm's supply curve and then showed how the horizontal summation of individual firms' supply curves generated a market supply curve. We cannot analyze long-run supply in the same way, however, because in the long run firms enter and exit the market as the market price changes. This makes it impossible to sum up supply curves-we don't know which firms' supplies to add.

To determine long-run supply, we assume all firms have access to the available production technology. Output is increased by using more inputs, not by invention. We also assume that the conditions underlying the market for inputs to production do not change when the industry expands or contracts. For example, an increased demand for labor does not increase a union's ability to negotiate a better wage contract for its workers.

The shape of the long-run supply curve depends on the extent to which increases and decreases in industry output affect the prices that the firms must pay for inputs into the production process. It is thus useful to distinguish among three types of industries: constant-cost, increasing-cost, and decreasing-cost

Constant-Cost Industry

Figure 8.15 shows the derivation of the long-run supply curve for a constant-cost industry. Assume that the industry is initially in long-run equilibrium at the intersection of market demand curve D_1 and market supply curve S_1, in part (b) of the figure. Point A at the intersection of demand and supply is on the long-run supply curve S_L because it tells us that the industry will produce Q_1 units of output when the long-run equilibrium price is P_1.

To obtain other points on the long-run supply curve, suppose the market demand for the product unexpectedly increases, say because of a tax cut. A typical firm is initially producing at an output of q_1, where P_1 is equal to long-run marginal cost and long-run average cost. But the firm is also in short-run equilibrium, so that price also equals short-run marginal cost. Suppose that

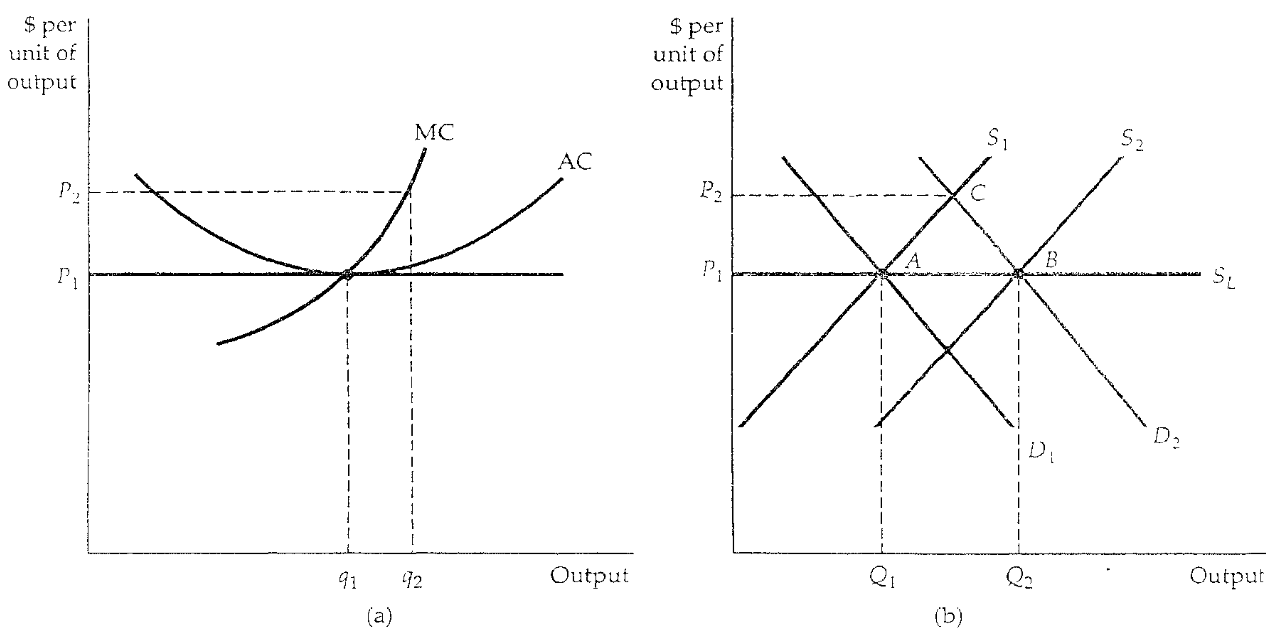

FIGURE 8.15 Long-Run Supply in a Constant-Cost Industry. In (b) the long-run supply curve in a constant-cost industry is a horizontal line S_L. When demand increases, initially causing a price rise, the firm. initially increases its output from q_1 to q_2 in (a). But the entry of new firms causes a shift to the right in supply. Because input prices are unaffected by the increased output of the industry, entry occurs until the original price is obtained.

the tax cut shifts the market demand curve from D_1 to D_2. Demand curve D_2 intersects supply curve Si at C As a result,the price increases from P_1 to P_2.

Part (a) shows how this price increase affects a typical firm in the industry. When the price increases to P_2, the firm follows its short-run marginal cost curve and increases its output to q_2. This output choice maximizes profit because it satisfies the condition that price equal short-run marginal cost. If every firm responds this way, each firm will be earning a positive profit in short-run equilibrium. This profit will be attractive to investors and will cause existing firms to expand their operations and new firms to enter the market.

As a result, in Figure 8.15b the short-run supply curve shifts to the right, from S_1 to S_2. This shift causes the market to move to a new long-run equilibrium at the intersection of D_2 and S_2. For this intersection to be a long-run equilibrium, output must expand just enough so that firms are earning zero profit and the incentive to enter or exit the industry disappears.

In a constant-cost industry, the additional inputs necessary to produce the higher output can be purchased without an increase in the per-unit price. This might happen, for example, if unskilled labor is a major input in production, and the market wage of unskilled labor is unaffected by the increase in the demand for labor. Since the prices of inputs have not changed, the firms' cost curves are also unchanged; the new equilibrium must be at a point such as *B* in Figure 8.15b, at which price is equal to Pi, the original price before the unexpected increase in demand occurred.

The long-run supply curvefor a constant-cost industry is, therefore, a horizontal line at a price that is equal to the long-run minimum average cost of production. At any higher price, there would be positive profit, increased entry, increased short-run supply, and thus downward pressure on price. Remember that in a constant-cost industry, input prices do not change when conditions change in the output market. Constant-cost industries can have horizontal long-run average cost curves.

Increasing-Cost Industry

In an increasing-cost industry, the prices of some or all inputs to production increase as the industry expands and the demand for the inputs grows. This might arise, for example, if the industry uses skilled labor, which becomes in short supply as the demand for it increases. Or the firm might require mineral resources that are available only *on* certain types of land, so that the cost of land as an input increases with output. Figure 8.16 shows the derivation of long-run supply, which is similar to the previous constant-cost derivation. The industry is initially in long-run equilibrium at A in part (b). When the demand curve unexpectedly shifts from D_1 to D_2, the short-run price of the product increases to P_2, and industry output increases from Q_1 to Q_2. A typical firm shown in part (a) increases its output from q_1 to q_2 m response to the higher price by moving along its short-run marginal cost curve. The higher profit that this and other firms earn induces new firms to enter the industry.

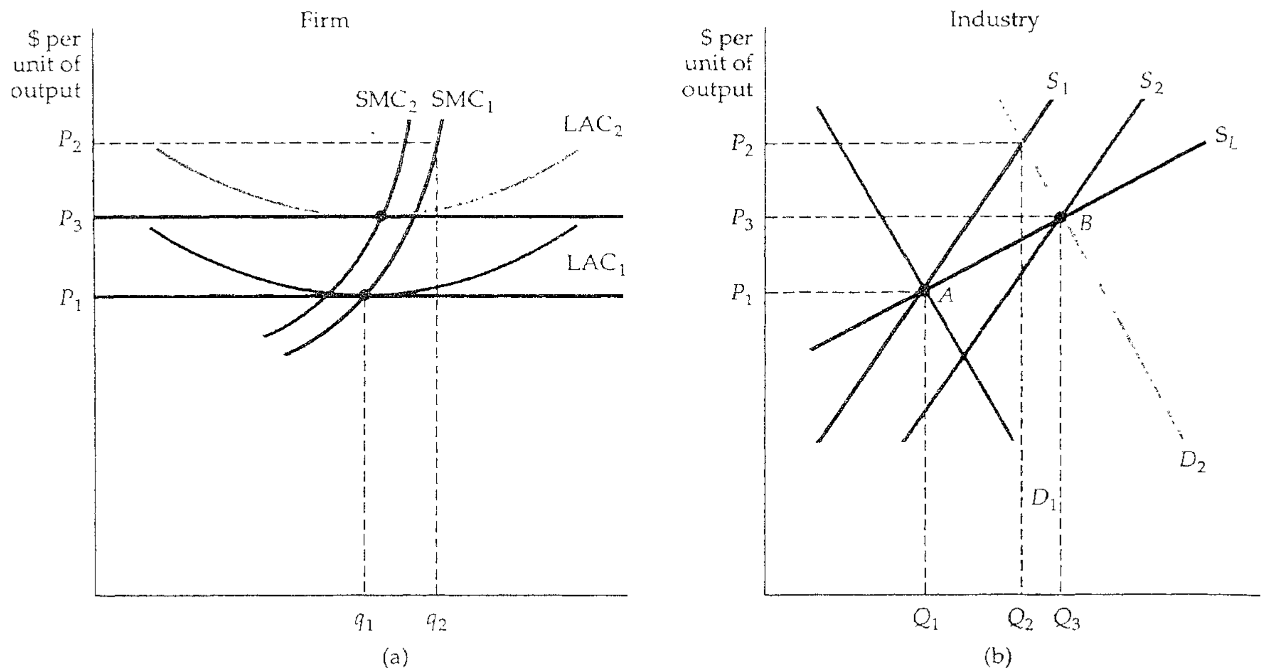

FIGURE 8.16 Long-Run Supply in an Increasing-Cost Industry. In (b), the long-run supply curve in an increasing-cost industry is an upward-sloping curve S_L. When demand increases, initially causing a price rise, the firms increase their output from q_1 to q_2 in (a). Then, the entry of new firms causes a shift to the right in supply. Because input prices increase as a result, the new long-run equilibrium occurs at a higher price than the initial equilibrium.

As new firms enter and output expands, the increased demand for inputs causes some or all input prices to increase. The short-run market supply curve shifts to the right as before, but not as much, and the new equilibrium at *B* results in a price P_3 that is higher than the initial price P_1. The higher market price is needed to ensure that firms earn zero profit in long-run equilibrium because the higher input prices raise the firms7 short-run and long-run cost curves. Figure 8.16a illustrates this. The long-run average cost curve shifts up from LAC_1 to LAC_2, while the short-run marginal cost curve shifts to the left from SMC_1 to SMC_2. The new long-run equilibrium price P_3 is equal to the new long-run minimum average cost. As in the constant-cost case, the higher short-run profit caused by the initial increase in demand disappears in the long run as firms increase their output and input costs rise.

The new long-run equilibrium at *B* in Figure 8.16b is, therefore, on the long-run supply curve for the industry. *In an increasing-cost industry, the long-run industry supply curve is upward sloping.* The industry produces more output, but only at the higher price needed to compensate for the increase in input costs. The term "increasing cost" refers to the upward shift in the firms' long-run average cost curves, not to the positive slope of the cost curve itself.

Decreasing-Cost Industry

The industry supply curve can also be downward sloping. In this case, the unexpected increase in demand causes industry output to expand as before. But as the industry grows larger, it can take advantage of its size to obtain some of its inputs more cheaply. For example, a larger industry may allow for an improved transportation system or for a better, less expensive financial network. In this case firms' average cost curves shift downward (even though firms do not enjoy economies of scale), and the market price of the product falls. The lower market price and the lower average cost of production induce a new long-run equilibrium with more firms, more output, and a lower price. Therefore, *in a decreasing-cost industry, the long-run supply curve for the industry is downward sloping.*

It is tempting to use the decreasing-cost argument to explain why computers have fallen in price over time. But other explanations are usually more persuasive. For example, lower computer prices can be explained by improvements in technology that lower production costs, or by a learning curve. The long-run, downward-sloping supply curve arises only when expansion itself lowers input prices, or when firms can use scale or scope economies to produce at lower cost.

The Short-Run and Long-Run Effects of a Tax

In Chapter 6 we saw that a tax on a firm's input (in the form of an effluent fee) creates an incentive for the firm to change the way it uses inputs in its production process. Now we consider how a firm responds to a tax on its output. To simplify the analysis, assume that the firm uses a fixed-proportions production technology. If the firm is a polluter, the output tax can encourage the firm to reduce its output, and therefore its effluent, or the tax might be imposed just to raise revenue.

First, suppose the output tax is imposed only on this firm, and thus does not affect the market price of the product. We will see that the tax on output encourages the firm to reduce its output. Figure 8.17 shows the relevant short-run cost curves for a firm enjoying positive economic profit by producing an output of q_1 and selling its product at the market price P_1. Because the tax is assessed for every unit of output, it raises the firm's marginal cost curve from MC_1 to $MC_2 = MC_1 + t$, where t is the tax per unit of the firm's output. The tax also raises the average variable cost curve by the amount t.

A close look at Figure 8.17 shows us that the output tax can have two possible effects. First, if the tax is less than the firm's profit margin, the firm will maximize its profit by choosing an output at which its marginal cost plus the

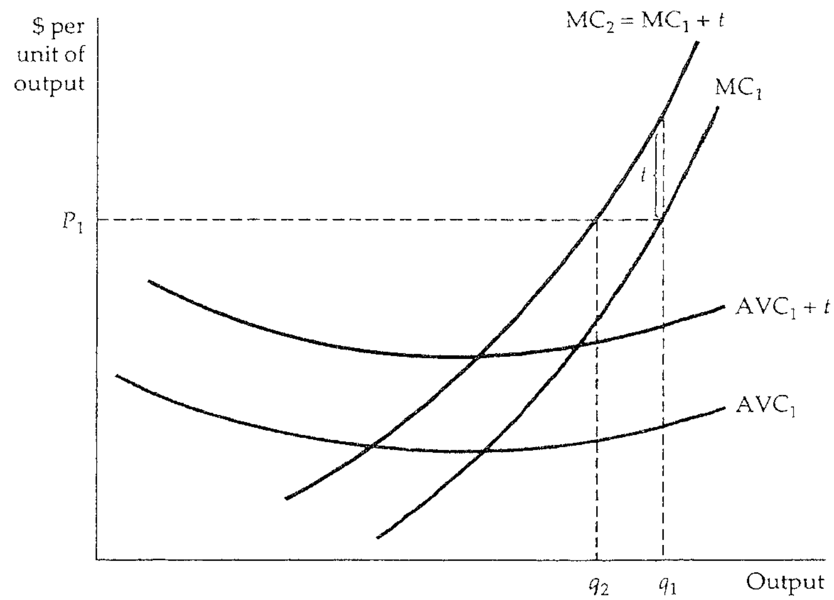

FIGURE 8.17 Effect of an Output Tax on a Competitive Firings Output. An output tax raises the firm's marginal cost curve by the amount of the tax. The firm will reduce its output to the point at which the marginal cost plus the tax is equal to the price of the product.

tax is equal to the price of the product. The firm's output falls from q_1 to q_2 and the implicit effect of the tax is to shift the firm's short-run supply curve upward (by the amount of the tax). Second, if the tax is greater than the firm's profit margin, then the average variable cost curve will rise, and the minimum average variable cost will be greater than the market price of the product. The firm will then choose not to produce.

Now suppose all firms in the industry are taxed and have increasing marginal costs. Since each firm reduces its output at the current market price, the total output supplied by the industry will also fall, causing the price of the product to increase. Figure 8.18 illustrates this, where an upward shift in the supply curve,from S_1 to $S_2 = S_1 + t$, causes the market price of the product to increase (by less than the amount of the tax) from P_1 to P_2 This increase in the price of the product diminishes some of the effects that we described previously. Firms will reduce their output less than they would without a price increase.

Output taxes may also encourage some firms (those whose costs are somewhat higher than others) to exit the industry. Figure 8.19 shows the long-run effects of the tax. Part (a) shows that the fee raises the long-run average cost curve for each firm. This makes production unprofitable for some firms, which choose to exit the industry in search of greater profit elsewhere. This results

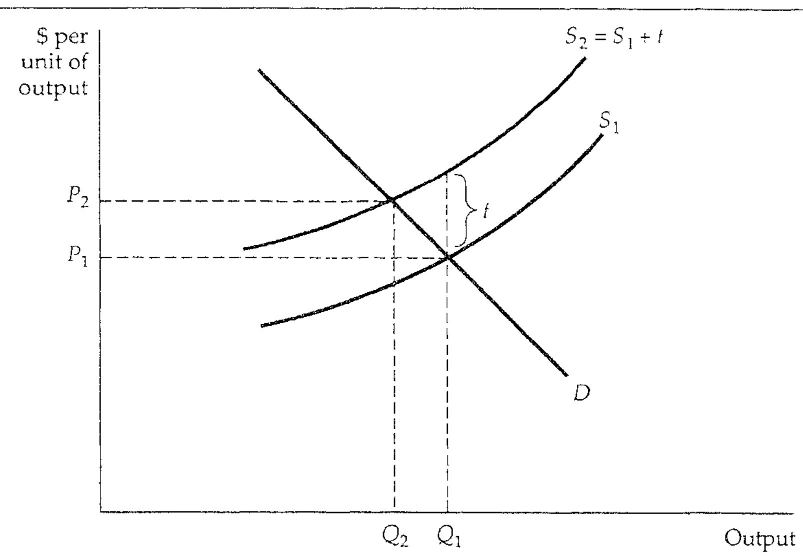

FIGURE 8.18 Effect of an Output Tax on Industry Output. An output tax placed on all firms in a competitive market shifts the short-run supply curve for the industry upward by the amount of the tax. This raises the market price of the product and lowers the total output of the industry.

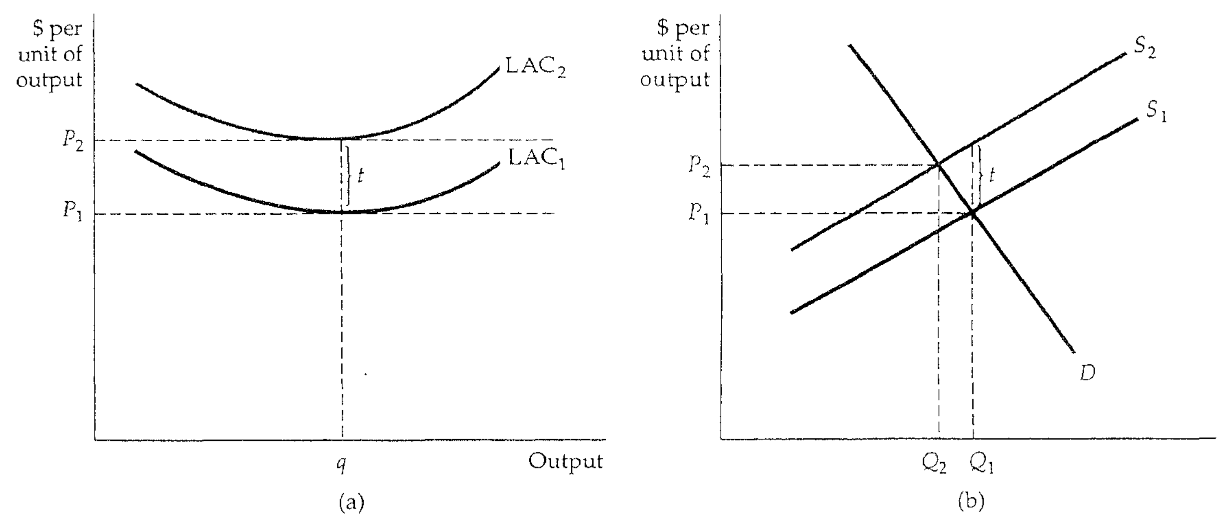

FIGURE 8.19 The Long-Run Effects of an Output Tax. In the long run, the output tax will raise the average cost curve in (a) from LAC_1 to LAC_2. As firms exit the industry because of lower profits, the aggregate supply curve in (b) shifts upward and to the left In the long-run equilibrium, quantity demanded and quantity supplied are equated at a higher price and a lower output.

in a shift to the left in the market supply curve, shown in part (b). The market price of the product increases from $P1$ to $P2$ and the quantity sold in the market falls from $Q1$ to $Q2$.

When the dust settles, the long-run equilibrium will have fewer firms and less output (and less effluent produced), because the output tax has reduced the relative profitability of production in the industry and has encouraged some investors to look elsewhere.[10]

Long-Run Elasticity of Supply

The long-run elasticity of industry supply is defined in the same way as the short-run elasticity. It is the percentage change in output $(\Delta Q/Q)$ that results from a percentage change in price $(\Delta P/P)$. a constant-cost industry, the long-run supply curve is horizontal, and the long-run supply elasticity is infinitely large. (A small increase in price will induce an extremely large increase in output.) In an increasing-cost industry, however, the long-run supply elasticity will be positive. Because industries can adjust and expand in the long run, we would generally expect long-run elasticities of supply to be larger than short-run elasticities." The magnitude of the elasticity will depend on the extent to which input costs increase as the market expands. For example, an industry that depends on inputs that are widely available will have a more elastic long-run supply than will an industry that uses inputs in short supply.

EXAMPLE 8.4 THE LONG-RUN SUPPLY OF HOUSING

Owner-occupied and rental housing provide interesting examples of the range of possible supply elasticities. People buy or rent housing to obtain the services that a house provides-a place to eat and sleep, comfort, and so on. If the price of housing services were to rise in one area of the country, the quantity of services provided could increase substantially.

To begin, consider the supply of owner-occupied housing in suburban or rural areas where land is not scarce. Here, the price of land does not increase substantially as the quantity of housing supplied increases. Likewise, the costs associated with construction are not likely to increase because there is a national market for lumber and other materials. Therefore, the long-run elasticity of the supply of housing is likely to be very large, approximating a constant-cost

[10] Although total market output will decline, each firm that remains in the market could produce more output and generate more effluent if the increase in the price is greater than the upward shift in the long-run average cost curve. But if policy is directed toward total industrial pollution/ the response of the market, not of individual firms, is important.

[11] In Some cases the opposite is true. Consider the elasticity of supply of scrap metal from a durable good like copper. Recall from Chapter 2 that because there is an existing stock of scrap, the long-run elasticity of supply will be *smaller* than the short-run elasticity.

industry. In fact, one recent study found the long-run supply curve to be nearly horizontal.[12]

Even when the elasticity of supply is measured within urban areas, where land costs rise as the demand for housing services increases, the long-run elasticity of supply is still likely to be large because land costs make up only about one-quarter of total housing costs. In one study of urban housing supply, the price elasticity was found to be 5.3.[13]

The market for rental housing is different, however. The construction of rental housing is often restricted by local zoning laws. Many communities outlaw it entirely, while others limit it to certain areas. Because urban land on which most rental housing is located is restricted and valuable, the long-run elasticity of supply of rental housing is much lower than the elasticity of supply of owner-occupied housing. As the price of rental housing services rises, new high-rise rental units are built, and older units are renovated, which increases the quantity of rental services. With urban land becoming more valuable as housing density increases, and with the cost of construction soaring with the height of buildings, the increased demand causes the input costs of rental housing to rise. In this increasing-cost case, the elasticity of supply can be much less than one; in one study the authors found the supply elasticity to be between 0.3 and 0.7.[14]

8.8 *When Is a Market Perfectly Competitiue?*

Apart from agriculture, few real-world markets are perfectly competitive in the sense that each firm faces a perfectly horizontal demand curve for a homogeneous product, and that firms can freely enter or exit the industry. Nevertheless, the analysis that we have just completed is useful because many markets are *almost* perfectly competitive: Firms in these markets face highly elastic demand curves, and entry and exit are relatively easy. As a result, it is profitable to set output so that the marginal cost of production is approximately equal to price.

A simple rule of thumb to describe whether a market is close to being perfectly competitive would be appealing. Unfortunately, we have no such rule, and it is important to understand why Consider the most obvious candidate:

[12] See James R. Follain, Jr., "The Price Elasticity of the Long-Run Supply of New Housing Construction,"*Land Economics* (May 1979): 190-199.

[13] See Barton A. Smith, "The Supply of Urban Housing,"*Journal of Political Economy* 40, No. 3 (Aug. 1976): 389-105.

[14] See Frank deLeeuw and Nkanta Ekanem, "The Supply of Rental Housing," *American Economic Review* 61 (Dec. 1971): 806-817, Table 5.2.

an industry with many *firms* (say at least 10 to 20). Unfortunately, the presence of many firms is neither necessary nor sufficient for an industry to approximate perfect competition because firms can implicitly or explicitly collude in setting prices.

The presence of only a few firms in a market also does not rule out competitive behavior. Suppose that five firms are in the market, but market demand for the product is very elastic. Then, the demand curve facing each firm is likely to be nearly horizontal, and the firms will behave *as if* they were operating in a perfectly competitive market. Or even if market demand is not very elastic, these five firms might compete aggressively (as we discuss in Chapter 13).

Contestable Markets

Even when only one firm is in a market, that firm may find it profit-maximizing to act as if it were competitive. The reason is that if it tries to raise price above the competitive level, other firms will enter the market, compete for customers, and force the price back down. Hence, competition among firms *within* a market can be less important than the competition/or *a market.* In a *contestable market,* new firms may enter the market under essentially the same. cost conditions as a firm that is already in the market. A firm can also exit the market without losing any investment in capital that is specific to that market and valueless elsewhere.[15]

Consider, for example, a neighborhood market for retail gasoline stations. It maybe economical for there to be only two stations in the neighborhood, since the presence of three or more would cause all stations to lose profits. Yet this market might be contestable, since the cost of opening a gas station is not *sunk,* i.e., it is not specific to that location. As a result, it is easy for firms to enter or exit the market as economic conditions change.

Most monopolistic or oligopolistic markets are not contestable, because the incumbent firms have sunk costs. These incumbents have a competitive advantage over prospective newcomers, and can charge a price higher than marginal cost. Suppose, for example, that a firm has a local monopoly over cable television. Economies of scale make it economical for one firm to provide cable service. But, the market is not fully contestable because some of the cost incurred by the cable company is sunk, and cannot be transferred if the company were to move its business elsewhere. The cable itself can be reutilized, but much of the labor involved in moving it would be wasted, and some of the cable and other materials associated with the hookups in each house would be valueless if the company had to exit the business. Because some of the investment is sunk, a new firm competing for the business would have to bid high enough to cover all its costs, whereas the incumbent firm could set a

[15] The theory is developed in William J. Baumol, John C. Panzar, and Robert D. Willig, *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt, Brace, Jovanovich, 1982), and criticized in William G. Shepherd, "Contestability vs. Competition," *American Economic Review* 74 (Sept. 1984): 572-587.

slightly lower price and make a substantial profit above and beyond its variable costs.

The point is that firms may behave competitively in many situations. Unfortunately, no simple indicator signifies when a market approximates perfect competition. Often it is necessary to analyze the number and size of films and their strategic interactions, as we do in Chapters 12 and 13.

Summary

1. The managers of firms can operate in accordance with a complex set of objectives and under various constraints. However, we can assume that firms act as if they are maximizing their long-run profit.
2. Because a firm in a competitive market has a small share of total industry output, it makes its output choice under the assumption that the demand for its own output is horizontal, in which case the demand curve and the marginal revenue curve are identical.
3. In the short run, a competitive firm maximizes its profit by choosing an output at which price is equal to (short-run) marginal cost, so long as price is greater than or equal to the firm's minimum average variable cost of production.
4. The short-run market supply curve is the horizontal summation of the supply curves of the firms in an industry. It can be characterized by the elasticity of supply—the percentage change in quantity supplied in response to a percentage change in price.
5. The producer surplus for a firm is the difference between revenue of a firm and the minimum cost that would be necessary to produce the profit-maximizing output. In both the short run and the long run, producer surplus is the area under the horizontal price line and above the marginal cost of production for the firm.
6. Economic rent is the payment for a scarce factor of production less the minimum amount necessary to hire that factor. In the long run in a competitive market, producer surplus is equal to the economic rent generated by all scarce factors of production.
7. In the long run, profit-maximizing competitive firms choose the output at which price is equal to long-run marginal cost.
8. A long-run competitive equilibrium occurs when (i) firms maximize profit; (ii) all firms earn zero economic profit, so that there is no incentive to enter or exit the industry; and (iii) the quantity of the product demanded is equal to the quantity supplied.
9. The long-run supply curve for a firm is horizontal when the industry is a constant-cost industry in which the increased demand for inputs to production (associated with an increased demand for the product) has no effect on the market price of the inputs. But the long-run supply curve for a firm is upward sloping in an increasing-cost industry, where the increased demand for inputs causes the market price of some 01 all inputs to production to rise.
10. Many markets may approximate perfect competition in that one or more firms act as if they face a nearly horizontal demand curve. However, the number of firms in an industry is not always a good indicator of the extent to which that industry is competitive.

Questions for Reuiew

1. Why would a *firm* that incurs losses choose to produce rather than shut down?

2. The supply curve for a firm in the short run is the short-run marginal cost curve (above the point of minimum average variable cost). Why is the supply curve in the long run *not* the long-run marginal cost curve (above the point of minimum average total cost)?

3. In long-run equilibrium, all firms in the industry earn zero economic profit. Why is this true?

4. What is the difference between economic profit and producer surplus?

5. Why do firms enter an industry when they know that in the long run economic profit will be zero?

6. At the beginning of the twentieth century, there were many small American automobile manufacturers. At the end of the century, there are only three large ones. Suppose that this situation is not the result of lax federal enforcement of antimonopoly laws. How do you explain the decrease in the number of manufacturers? (Hint: What is the inherent cost structure of the automobile industry?)

7. Industry X is characterized by perfect competition, so that every firm in the industry is earning zero economic profit. If the product price fell/ no firms could survive. Do you agree or disagree? Discuss.

8. An increase in the demand for video films also increases the salaries of actors and actresses. Is the long-run supply curve for films likely to be horizontal or upward sloping? Explain.

9. True or false: A firm should always produce at an output at which long-run average cost is minimized. Explain.

10. Can there be constant returns to scale in an industry with an upward-sloping supply curve? Explain.

11. What assumptions are necessary for a market to be perfectly competitive? In light of what you have learned in this chapter, why is each of these assumptions important?

12. Suppose a competitive industry faces an increase in demand (i.e., the curve shifts upward). What are the steps by which a competitive market insures increased output? Does your answer change if the government imposes a price ceiling?

13. The government passes a law that allows a substantial subsidy for every acre of land used to grow tobacco. How does this program affect the long-run supply curve for tobacco?

Exercises

1. From the data in Table 8.2, show what happens to the firm's output choice and profit if the price of the product falls from $40 to $35.

2. Again, from the data in Table 8.2, show what happens to the firm's output choice and profit if the fixed cost of production increases from $50 to $100, and then to $150.-What general conclusion can you reach about the effects of fixed costs on the firm's output choice?

3. Suppose you are the manager of a watchmaking firm operating in a competitive market. Your cost of production is given by $C = 100 + Q_2$, where Q is the level of output and C is total cost. (The marginal cost of production is $2Q$. The fixed cost of production is $100.)

- **a.** If the price of watches is $60, how many watches should you produce to maximize profit?
- **b.** What will the profit level be?
- **c.** At what minimum price will the firm produce a positive output?

4. Use the same information as in Exercise 1 to answer the following.

TABLE 8.2 A Firm's Short-Run Revenues and Costs

Output (Units)	Price ($/Unit)	Revenue ($)	Total Cost ($)	Profit ($)	Marginal Cost ($)	Marginal Revenue ($)
0	40	0	50	−50	—	—
1	40	40	100	−60	50	40
2	40	80	128	−48	28	40
3	40	120	148	−28	20	40
4	40	160	162	−2	14	40
5	40	200	180	20	18	40
6	40	240	200	40	20	40
7	40	280	222	58	22	40
8	40	320	260	60	38	40
9	40	360	305	55	45	40
10	40	400	360	40	55	40
11	40	440	425	15	65	40

a. Derive the firm's short-run supply curve. (Hint: You may want to plot the appropriate cost curves.)
b. If 100 identical firms are in the market, what is the industry supply curve?

5. A sales tax of $1 per unit of output is placed on one firm whose product sells for $5 in a competitive industry.
a. How will this tax affect the cost curves for the firm?
b. What will happen to the firm's price, output, and profit in the short, run?
c. What will happen in the long run?

6. Suppose that a competitive firm's marginal cost of producing output q is given by $MC(q) = 3 + 2q$. If the market price of the firm's product is $9:
a. What level of output will the firm produce?
b. What is the firm's producer surplus?

7. Suppose that the average variable cost of the firm inproblem 6 is given by $AVC(q) = 3 + q$. Suppose that the firm's fixed costs are known to be $3. Will the firm be earning a positive, negative, or zero profit in the short run?

8. A competitive industry is in long-run equilibrium. A sales tax is then placed on all firms in the industry. What do you expect to happen to the price of the product, the number of firms in the industry, and the output of each firm in the long run?

***9.** A sales tax of 10 percent is placed on half the firms (the polluters) in a competitive industry The revenue is paid to the remaining firms (the nonpolluters) as a 10 percent subsidy on the value of output sold.
a. Assuming that all firms have identical constant long-run average costs before the sales tax-subsidy policy, what do you expect to happen to the price of the product, the output of each of the firms, and industry output, in the short run and the long run? (Hint: How does price relate to industry input?)
b. Can such a policy *always* be achieved with a balanced budget in which tax revenues are equal to subsidy payments? Why? Explain.