Chapter 5
Perfect Competition and Market Imperfections*

Competition is a fundamental concept in a market economy. We can think of competition as firm rivalry, where one firm battles to gain a strategic advantage over its competitors. For example, General Motors and Ford have been competing with one another for over a century to produce better cars at lower cost and to create more catchy marketing campaigns. We can also think of competition as a type of market structure. Both concepts are important in industrial organization. In later chapters, we analyze various forms of competitive behavior. In this chapter, we review the market structure of perfect competition.

Before we begin, it is important to distinguish between a market and an industry. A market is a collection of all buyers and sellers, with sellers supplying substitutable goods to the same potential buyers. An industry generally ignores the buyer side of the market, referring to the collection of firms that sell substitutable goods. Market and industry are frequently used interchangeably, but there will be times when it is important to distinguish between them. Practical issues involving how to accurately define a relevant market will be taken up in Chap. 8.

5.1 The Assumptions of Perfect Competition

Although no real-world market is perfectly competitive, the model of perfect competition provides us with a benchmark of market efficiency from which real markets can be judged. In addition, the model is based on a set of assumptions that produces a simple explanation of price and output determination in a free market. In physics, assuming that we live in a perfect vacuum makes the study of how objects fall to earth much easier, even though perfect vacuums do not exist.

* This is a review chapter. You can learn more about the basic economic models discussed in the chapter from any introductory or intermediate microeconomics textbook, such as Frank and Bernanke (2008), Mankiw (2011), Bernheim and Whinston (2008), Pindyck and Rubenfield (2009), and Varian (2010). For more advanced treatments, see Nicholson and Snyder (2012) and Mas-Colell et al. (1995).

V.J. Tremblay and C.H. Tremblay, *New Perspectives on Industrial Organization*, Springer Texts in Business and Economics, DOI 10.1007/978-1-4614-3241-8_5,

Similarly, assuming perfect competition simplifies the study of markets. The model is highly reductionist but still provides a viable representation of some real-world markets, especially for agricultural commodities.

The basic demand and supply model derives from a market that is assumed to be perfectly competitive. For a market to be **perfectly competitive**, it must meet the following five conditions:

1. Firms are *profit maximizers*.[1]
2. Firms produce *perfectly homogeneous goods*. That is, there are no real or perceived differences between the products of different firms.
3. There are *many identical firms* in the industry.[2] Extreme models consider the case where the number of firms is close to infinity. In such a market, each firm is so small that its production has no effect on the equilibrium price. Because products are homogeneous and firms are small, firms are said to be price takers. This means that the market price is exogenous (i.e., taken as given) to the firm. A firm is unable to profitably raise its price above the market equilibrium price. The assumption that firms are identical implies that their cost functions are the same.
4. There are *no barriers to entry or exit*. This means that entrepreneurs can enter or exit a market at zero cost.
5. There are *no frictions or other forms of market imperfections*. No frictions means that all buyers and sellers are perfectly informed about market conditions, and transaction costs are zero. In addition, goods are private (not public goods), and there are no externalities. Firms do not impose any benefits or costs on others without compensation.

You can see why this is called "perfect" competition. Each of the assumptions characterizes an extreme or limiting case. Products are not just similar but identical. The number of firms approaches infinity. The cost of entry and exit is zero. And there are no frictions or externalities. In such a market, the price represents the true marginal cost to society of production. One of the goals of the book is to see how market outcomes change when one or more of these assumptions are violated. Before modifying them, we first need to understand the perfectly competitive model.

5.2 Firm Behavior in Perfect Competition

5.2.1 Firm Demand and Revenue Functions

Because a perfectly competitive firm is a price taker, its demand function is a horizontal line at the equilibrium price. The market equilibrium price, p^*, is determined by

[1] In Chap. 2 we discuss alternative firm motives.

[2] The model also assumes that there are many buyers. Unless otherwise indicated, we will assume many buyers throughout the book.

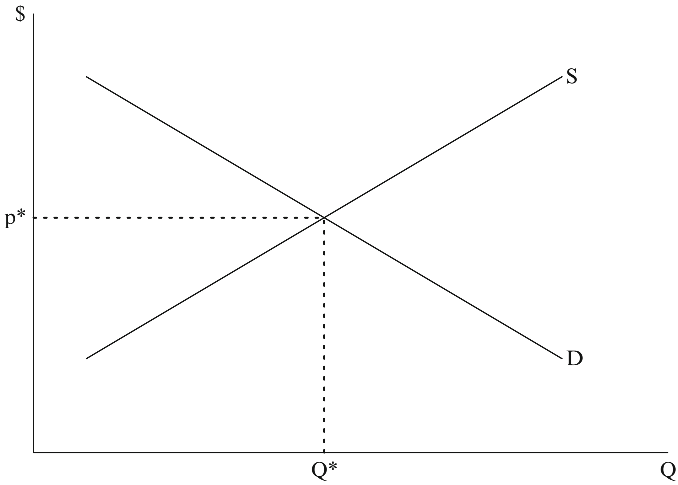

Fig. 5.1 Industry supply and demand

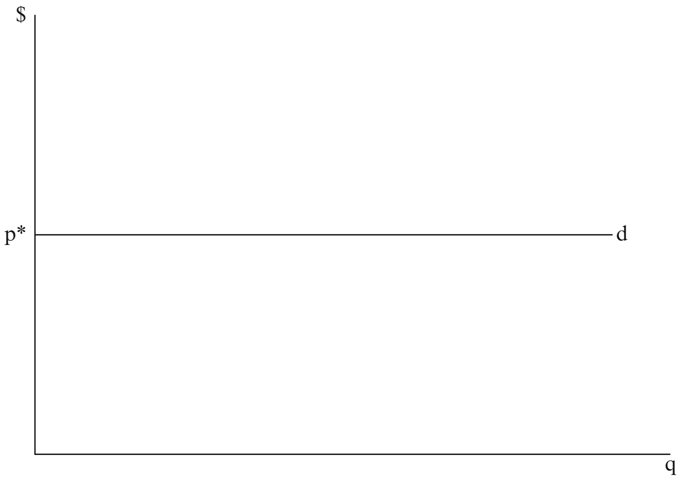

Fig. 5.2 The firm demand curve

industry supply (S) and market demand (D) as shown in Fig. 5.1, where Q^* is the equilibrium level of consumption in the market. If a firm charges a price above p^*, no consumers will buy from the firm because they can buy all they want at the equilibrium price. Thus, in perfect competition, a firm cannot charge a price higher than the market price. A firm will not sell at a price below p^*, because it can sell all that it is capable of producing at the equilibrium price. To lower price would lower profits unnecessarily. Under these conditions, firm demand is perfectly or infinitely elastic and represented by the horizontal line d in Fig. 5.2, where q is the firm's level of production.

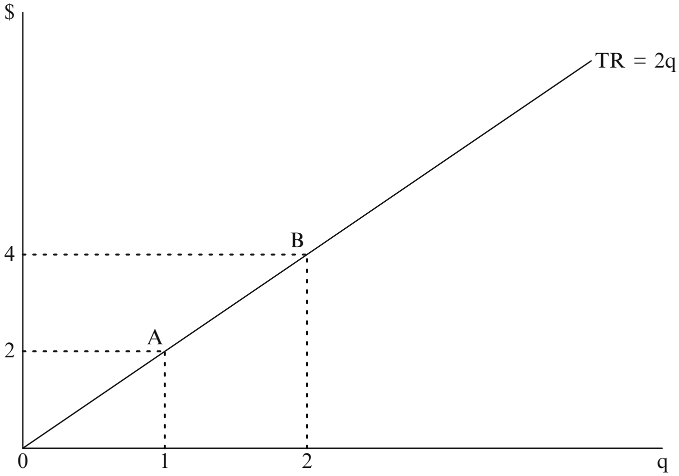

Fig. 5.3 The total revenue function when price = 2

Before discussing firm profit, we need to describe **total revenue** (TR) for the perfectly competitive firm. Generally, total revenue is price times quantity, TR = $p(q)q$, where $p(q)$ is inverse demand. Since a perfectly competitive firm will charge the same price regardless of the level of output, the total revenue function for the firm is

$$TR = pq. \tag{5.1}$$

Figure 5.3 shows the total revenue function for a hypothetical firm with a price of \$2, so that TR = $2q$. For 1 unit of output, TR is \$2 (point A). If $q = 2$, then TR = 2 × 2 = \$4 (point B), and so on. The TR function is a straight line from the origin, and the slope of TR = 2 in this example. In general, the TR function for the perfectly competitive firm is a line from the origin with slope equal to p.

Recall from Chap. 2 that **marginal revenue** is the additional revenue the firm receives from a small increase in production, and that the marginal revenue function is the slope of the total revenue function. The equation for marginal revenue is[3]:

$$MR = \frac{\partial TR}{\partial q}. \tag{5.2}$$

Since TR = pq in perfect competition and $p = p^*$,

$$MR = \frac{\partial (p^* q)}{\partial q} = p^*. \tag{5.3}$$

[3] Although q is the only variable in this example, we use ∂ instead of d to remind us that there are many other variables that are implicitly assumed to be held fixed.

Marginal revenue for a perfectly competitive firm is the market price because an additional unit of output brings p^* to the firm.

Average revenue is revenue per unit of output. The AR function is

$$\mathrm{AR} = \frac{\mathrm{TR}}{q}. \tag{5.4}$$

In perfect competition,

$$\mathrm{AR} = \frac{p^* q}{q} = p^*. \tag{5.5}$$

Thus, for a perfectly competitive firm,

$$\mathrm{MR} = \mathrm{AR} = p^*. \tag{5.6}$$

Geometrically, the MR, AR, and firm demand curve coincide. In reference to Fig. 5.2, the MR and AR functions are the same as demand, d.

5.2.2 *Profit Maximization*

How do firms decide how much output to produce when there is perfect competition? Firms are assumed to choose the amount of output that maximizes profit. Profit is defined as total revenue minus total cost (TC):[4]

$$\pi = \mathrm{TR} - \mathrm{TC}. \tag{5.7}$$

In general, the profit equation will be concave or hill-shaped, as described in Fig. 5.4. When $q = 0$, no output is produced and profit is 0. From $q = 0$ to $q = q^*$, profit is increasing at a decreasing rate. Profit peaks at q^*, decreases at a decreasing rate until $\pi = 0$ at q', where TR $=$ TC. Beyond q', TC $>$ TR and profit is negative. The firm will produce q^* which generates the highest level of profit. Notice that at the profit-maximizing output level, the slope of the tangent line to the profit equation is zero.

Another way to find q^* is with the use of calculus. To find the output level that maximizes profit, we take the derivative of π with respect to q and set it equal to zero:

$$\begin{aligned}\frac{\partial \pi}{\partial q} &= \frac{\partial \mathrm{TR}}{\partial q} - \frac{\partial \mathrm{TC}}{\partial q}, \\ &= \mathrm{MR} - \mathrm{MC} = 0,\end{aligned} \tag{5.8}$$

[4] Generally, we will consider long-run problems, but at this point, we are abstracting from time.

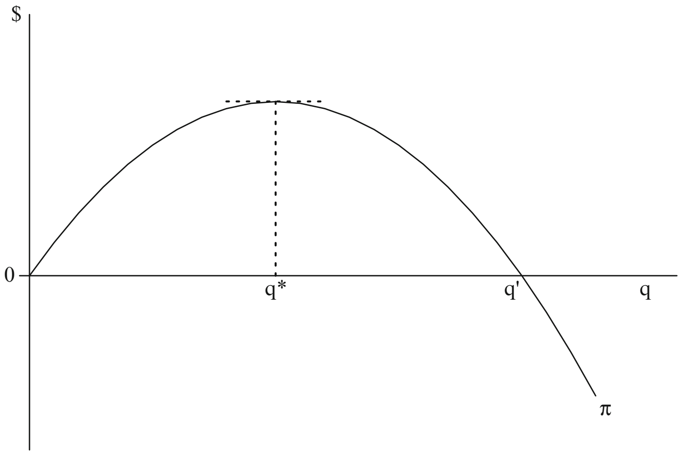

Fig. 5.4 The profit function

where MC is marginal cost (∂TC/∂q). Equation (5.8) is called the first-order condition of profit maximization. This equation produces the standard *marginal principle* or *rule of marginalism*: to maximize payoff with respect to activity x (output in this case), the firm must equate the marginal benefit (MR) with the marginal cost (MC) of activity x. This condition indicates that the firm will choose the output level where MR = MC.[5] Given that the market equilibrium price is identical to marginal revenue, the first-order condition for a perfectly competitive firm also implies that

$$p \equiv \mathrm{MR} = \mathrm{MC} \tag{5.9}$$

at the profit-maximizing level of output.

Referring back to (5.8), consider a simple example where MR $= 2$ and MC $= 2q$.[6] In this case,

$$\begin{aligned} \frac{\partial \pi}{\partial q} &= \mathrm{MR} - \mathrm{MC}, \\ &= 2 - 2q = 0. \end{aligned} \tag{5.10}$$

[5] A maximum is assured because the profit equation is concave. If it were convex, this method would identify the output level that would minimize profit. See the Mathematics and Econometrics Appendix at the end of the book for more details.

[6] These functions derive from the following total revenue, total cost, and profit functions: $TR = 2q$, $TC = q^2$, and $\pi = TR - TC = 2q - q^2$. The TC equation is not the usual representation of a cost equation in perfect competition, as we will see subsequently, but we use it here to provide a simple example.

Solving for the optimal level of output yields:

$$q^* = 1. \tag{5.11}$$

These general methods identify the firm's profit-maximizing output level, whether we are discussing the short run or the long run. As you recall, in the short run some inputs are fixed whereas in the long run all inputs are variable. For the remainder of the chapter and for most of the book, we focus on the long run.

Unfortunately, this discussion of firm behavior is incomplete, as it ignores the firm's **participation constraint**. The participation constraint identifies conditions under which the firm will stay in business versus shut down. That is, there are situations where it is optimal for the firm to produce zero output, even though the first-order condition indicates otherwise. In the example above, the firm is making an economic profit: $\pi^* = 1$, when $q^* = 1$. Thus, the firm has an incentive to stay in business. If, however, profits were negative, the firm would go out of business or shut down in the long run.[7] The long-run participation constraint is that the firm earns nonnegative profits. Thus, to fully understand firm behavior we must analyze both the firm's first-order condition and participation constraint.

5.3 Market Equilibrium and Long-Run Supply

In perfect competition, three conditions will hold when the market is in **long-run equilibrium** (i.e., there is no incentive for change):

- The market will clear. This means that demand equals long-run supply.
- Given profit maximization, the equilibrium price will equal long-run marginal cost, $p^* = \text{MC}$.
- Firm (economic) profit will be zero. The free entry/exit assumption assures this zero-profit condition. If profits are positive, firms will enter the market. Entry causes industry supply to increase, equilibrium price to fall, and profits to fall. This process will continue until profits equal zero. Alternatively, if profits are negative, firms will exit the market according to the long-run shutdown condition of profit maximization. This will raise profits and continue until profits equal zero. Profits equal zero when the equilibrium price equals long-run average cost (AC), $p^* = \text{AC}$.

This outcome is illustrated in Fig. 5.5, where the left-hand figure identifies firm costs and the right-hand figure identifies market demand (D) and long-run supply (S) conditions. We can summarize these conditions as follows: $\text{D} = S$ and $p^* = \text{MC} = \text{AC}$.

[7] In the short run, the firm must pay its fixed costs whether it shuts down or not. In this case, the firm would shut down if its losses from staying in business were less than its fixed cost, which it cannot avoid.

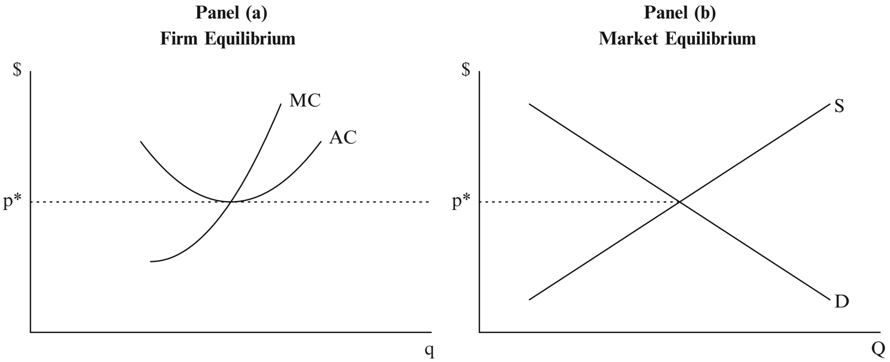

Fig. 5.5 Long-run equilibrium in perfect competition

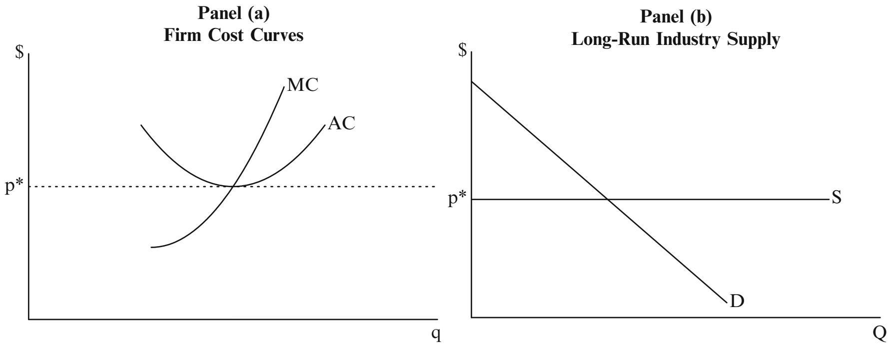

Fig. 5.6 Long-run supply in a constant-cost industry

Because our focus will be primarily on long-run outcomes, it will be useful to review the **long-run industry supply function**.[8] There are three cases, but we will focus only on the two cases that we will use in the book: the constant-cost case and the increasing-cost case.[9]

The **constant-cost** case assumes that the industry is so small that an increase or decrease in industry production (Q) has no effect on input prices, such as the price of labor, capital, and materials. In this case, AC and MC in Fig. 5.5 remain constant as industry production changes. As a result, the long-run equilibrium price is fixed at the minimum AC, and the long-run industry supply curve (S) is a horizontal line at that point, as illustrated in Fig. 5.6. The constant-cost case has an interesting

[8] This differs from the short-run supply function. Recall from principles of economics that a firm's short-run supply curve is its marginal cost curve above average variable cost. The short-run industry supply curve is the (horizontal) summation of the marginal cost curves of every firm in the industry. Thus, it reflects the industry's marginal cost of production.

[9] The third is the decreasing-cost case.

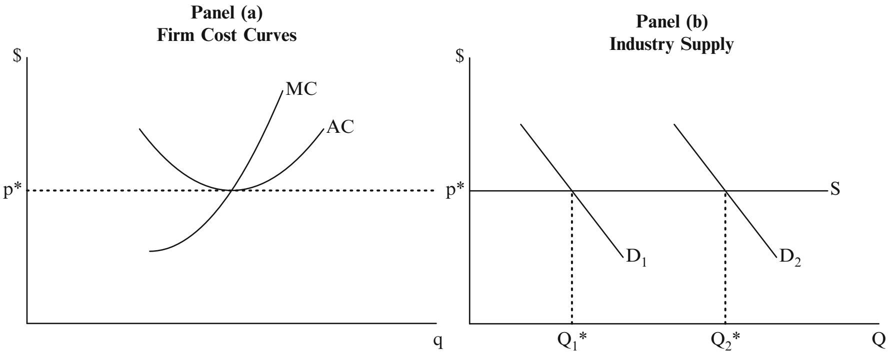

Fig. 5.7 A demand increase in a constant-cost industry

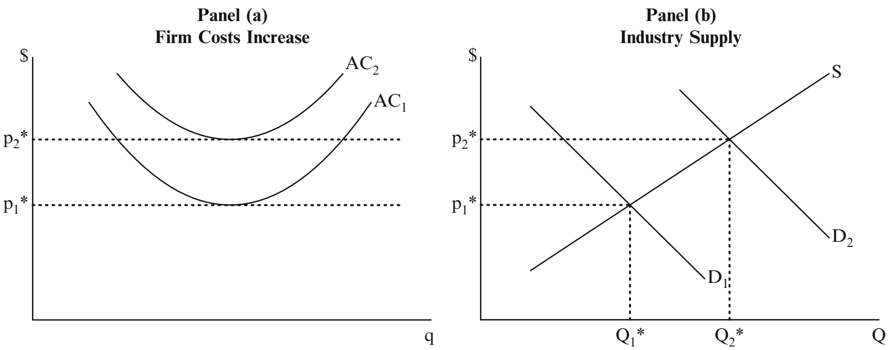

Fig. 5.8 Long-run supply in an increasing-cost industry

feature. The equilibrium price is determined by technology (or the minimum point on AC), while demand determines the equilibrium level of output. Demand has no effect on the long-run equilibrium price. Figure 5.7 illustrates this result for an increase in demand from D_1 to D_2 that causes the equilibrium quantity to change from Q_1^* to Q_2^* but has no effect on p^*.

The **increasing-cost** case produces the traditional long-run supply function that has a positive slope. In this case, an increase in industry production leads to an increase in input prices. This causes the minimum AC point to shift up and p^* to increase as industry production increases. Figure 5.8 shows that when demand increases from D_1 to D_2: (1) equilibrium quantity increases from Q_1^* to Q_2^*, (2) AC rises from AC_1 to AC_2,[10] and (3) the equilibrium price increases from p_1^* to p_2^*. Long-run supply is identified by the equilibrium price and equilibrium output

[10] To see the result more clearly, we leave off the MC_1 and MC_2 curves.

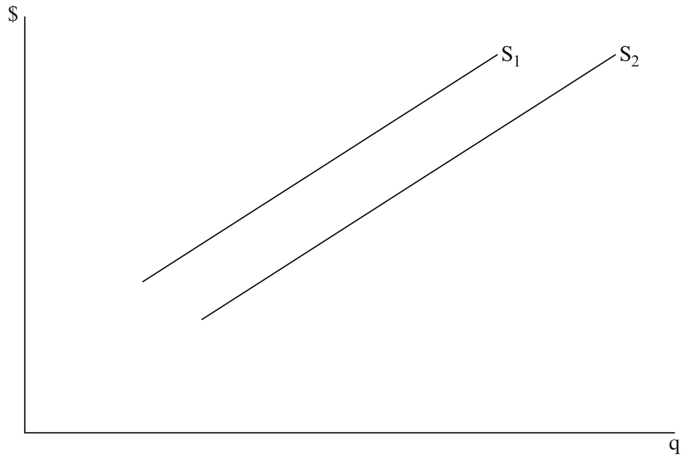

Fig. 5.9 An efficiency-enhancing technological change and firm supply

points. Thus, the long-run supply function has the usual positive slope.[11] Note that long-run supply always reflects the long-run cost of production, because the supply price will equal the minimum AC which equals MC.

Because the long-run supply function depends on the marginal cost of production, supply determinants are the same as cost determinants. That is, long-run industry supply is a function of the price, input prices, and technology. A change in the output price causes a movement along the supply function (in the increasing-cost case). A technological change or an exogenous decrease in the price of one or more inputs causes costs to fall and supply to shift down and to the right, such as in Fig. 5.9, where supply shifts from S_1 to S_2. One way of thinking about this change is that it allows the industry to profitably produce more output at the same price when there are lower production costs. Thus, it represents an increase in supply. The supply curve shifts up and to the left for a decrease in supply.

A useful tool in supply analysis is the **price-elasticity of supply** (ϵ_s), defined as the percentage change in quantity supplied (Q_s) resulting from a percentage change in price. For large changes it is defined as $(\Delta Q_s/Q_s)/(\Delta p/p)$. At times, we are interested in the effect on supply of a small change in price, which requires the use of calculus. In this case,

$$\epsilon_s \equiv \frac{\partial Q_s/Q_s}{\partial p/p} = \frac{\partial Q_s}{\partial p}\frac{p}{Q_s}. \tag{5.12}$$

[11] The third case involves decreasing costs. That is, input prices fall as industry production increases. The usual example given to justify this possibility is the presence of economies of scale in the production of a primary input. Scale economies cause the input price to fall with increased production, and the long-run supply function has a negative slope.

The price elasticity conveys the direction and magnitude of change in quantity supplied due to a change in the price. As with the demand curve, the price elasticity of supply can vary along a given supply curve.

The sign of the long-run supply elasticity depends on the cost conditions of the industry. For an increasing-cost industry, supply slopes upwards ($\partial Q_s/\partial p > 0$) and $\epsilon_s > 0$. For a constant-cost industry, supply is horizontal and price supply does not change with output. Thus, $\partial Q_s/\partial p = \infty$ and supply is said to be perfectly or infinitely elastic.[12] If supply is fixed at a particular level of Q, $\partial Q_s/\partial p = 0$ and $\epsilon_s = 0$; supply is perfectly inelastic.[13]

The extent of the long-run response of Q_s to a price change depends on the degree of the impact on input prices. In the case of an increasing-cost industry, if input prices rise substantially as quantity supplied increases, the supply elasticity will be greater than if input prices rise only slightly.

5.4 Comparative Statics

We can perform comparative static analysis on equilibrium models such as demand and supply. **Comparative statics** is the analysis of the change in the endogenous variables of a model that results from a change in an exogenous variable. We call this comparative statics because it involves a comparison of two "static" equilibria, ignoring the process that gets us from one equilibrium to another.

To illustrate, consider a simple demand and supply model in an increasing cost industry as depicted in Fig. 5.10. The market is initially in equilibrium at point *A* where demand (D_1) equals supply (S). We know that the demand for normal good x will increase with an increase in income, an increase in the price of a substitute, or a decrease in the price of a complement. One or more of these changes will cause demand to shift right from D_1 to D_2. Given sufficient adjustment time, this increase in demand will cause the equilibrium to change from point *A* to point *B*, causing an increase in both the equilibrium price (from p_1^* to p_2^*) and equilibrium quantity (from Q_1^* to Q_2^*).

Alternatively, a decrease in the price of an important input or a technological change will cause supply to rise, such as from S_1 to S_2 in Fig. 5.11. The equilibrium will change from point *A* to point *B*, causing equilibrium price to fall (from p_1^* to p_2^*) and equilibrium quantity to rise (from Q_1^* to Q_2^*).

We provide a more detailed discussion of comparative static analysis in the Appendix at the end of the book.

[12] For a decreasing-cost industry, $\partial Q_s/\partial p < 0$ and $\epsilon_s < 0$.

[13] This would be true, for example, for commodities that are not reproducible, such as van Gogh paintings.

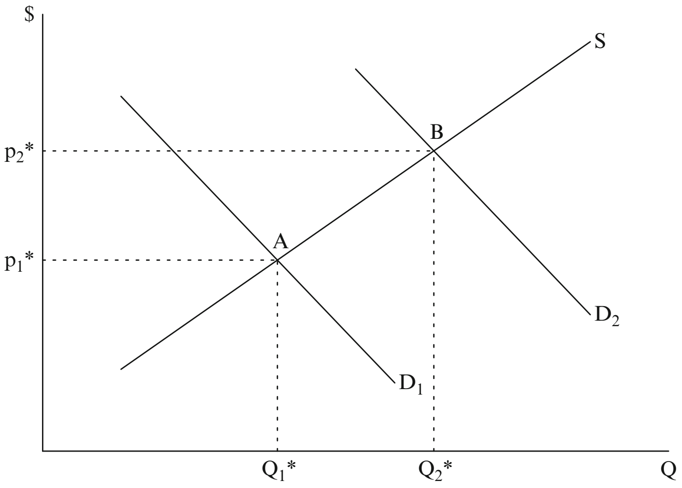

Fig. 5.10 An increase in the demand for good *x*

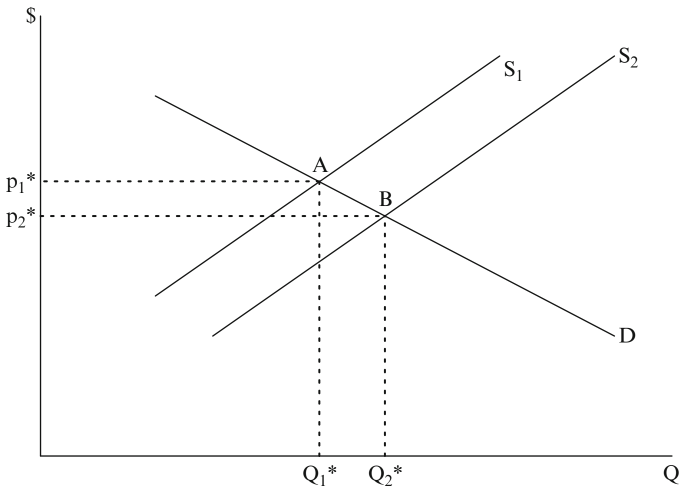

Fig. 5.11 An increase in the supply of good *x*

5.5 Efficiency and Welfare

In Chap. 1 we saw that desirable market performance requires that four conditions be met. A market outcome must be statically efficient, dynamically efficient, and equitable. When these, along with macro stability, are met in every market, social welfare is maximized. Dynamic efficiency issues will be discussed in Chap. 17, and equity issues will be discussed in Chap. 19. In this section, we are interested in static efficiency.

The concept of **static efficiency** is crucial to the study of welfare economics. Static efficiency is divided into four types. Two apply specifically to the firm:

- **Technical Efficiency**: A firm is technically efficient when it uses the minimum quantity of inputs to produce a given output. In other words, inputs are not being wasted.[14]
- **Economic Efficiency**: A firm is economically efficient when it produces a given output at minimum cost. Inputs are not being wasted and the firm is using the lowest cost combination of inputs to produce a given output.[15] When economically efficient, the firm is producing on (not above) its cost function.

Note that economic efficiency implies technical efficiency, but technical efficiency need not imply economic efficiency. For example, both rickshaw and auto taxies may be technically efficient modes of travel for short distances in large cities. They simply use different combinations of inputs to provide taxi service—an auto taxi uses gas but takes less time (i.e., uses fewer labor hours); a rickshaw uses no gas but takes considerable time. In less developed countries where gas is expensive and wages are low, rickshaw taxies are economically efficient. In developed countries where wages are relatively high, auto taxies are economically efficient.

There are also two types of efficiency at the industry level:

- **Productive Efficiency**: An industry is productively efficient when it produces a given level of output at minimum cost.
- **Allocative Efficiency**: An industry is allocatively efficient when it produces the socially desirable quantity of output. This occurs when the marginal benefit to society of producing another unit of output, reflected in the market price, equals the marginal cost of production.

You might think that if every firm is economically efficient, the industry must be productively efficient, but this is incorrect. For instance, consider a market where the long-run average cost function declines up to where it intersects the market demand function. In this case, a market with many small producers that are economically efficient (i.e., where each firm minimizes its cost of production) does not produce a productively efficient outcome. Industry costs are minimized with just one firm. This represents the natural monopoly case, which we discuss in the next chapter.

Baumol et al. (1982) discuss the formal link between productive efficiency and market structure. They developed the concept of a **cost-minimizing industry structure**, defined as the number of firms that minimizes industry production

[14] This means that the firm is operating on its isoquant, where an isoquant represents all minimum combinations of inputs that can just produce a given level of output. For further discussion, see Färe et al. (1985) and Varian (2010).

[15] The economically efficient point occurs where the firm's isoquant is tangent to its isocost function. The isocost function represents all combinations of inputs that generate a given cost. For further discussion, see Färe et al. (1985) and Varian (2010).

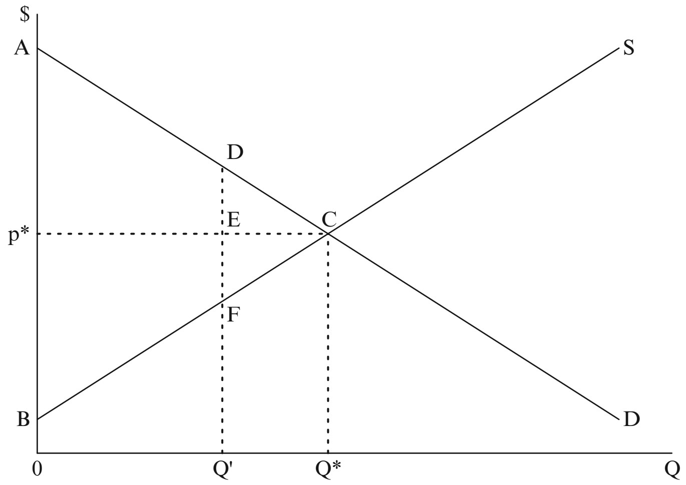

Fig. 5.12 Consumer surplus, producer surplus, and total surplus

costs. By comparing the cost minimizing structure to the actual structure of an industry, we are able to determine whether or not an industry is productively efficient. We will discuss this topic further in Chap. 8.

The reason that the perfectly competitive model is considered a welfare benchmark is that it produces a perfectly efficient outcome in the long run.[16] First, every firm must be economically efficient (and therefore technically efficient) because firms are assumed to be profit maximizers. To maximize profits, the firm must choose the combination of inputs that minimizes its cost of production. Second, the industry is productively efficient in the long run, because every firm produces at the minimum of its long-run average cost function.

Finally, we examine whether the long-run equilibrium in the perfectly competitive model is allocatively efficient by considering the concepts of consumer surplus, producer surplus, and total surplus.[17] Sometimes consumers pay less than they are willing to pay (i.e., the demand price), ending up with a net gain in welfare. **Consumer surplus** (CS) is the net gain that consumers receive from consuming a given quantity of output for a particular price. In other words, it is the difference between what a consumer is willing to pay for a good and what he or she actually pays. Graphically, consumer surplus is the area under the demand curve and above the price line for a given level of output. In Fig. 5.12, when price is p^* and quantity is Q', consumer surplus is area Ap^*ED.

[16] Here, we mean that the perfectly competitive model is statically efficient, but it need not be dynamicaly efficient as we will see in later chapters.

[17] This is a partial equilibrium approach, which ignores the effect that a price change in this market will have on other markets. For discussion of these general equilibrium effects, see Bernheim and Whinston (2008), Pindyck and Rubenfield (2009), and Varian (2010).

Just as consumers can receive a surplus, so can producers. Some may be willing to sell their product for less than the market price. **Producer surplus** (PS) is the difference between the price that producers receive for goods sold and the opportunity cost of all inputs used to produce output (i.e., the supply price, which equals marginal cost and is represented by the supply function). Producer surplus for all producers can be depicted graphically as the area under the price line and above the supply curve for a given quantity of output. Returning to Fig. 5.12, this is area p^*BFE at p^* and Q'. [18] Producer surplus represents producer welfare.

Total surplus (TS) is the sum of consumer surplus and producer surplus (TS = CS + PS) and measures total social welfare in this market. Total surplus is the area under the demand curve and above the supply curve, area ABFD at p^* and Q' in Fig. 5.12. Total surplus is the difference between total benefits derived from the good (area under demand curve) and total costs of producing the good (area under supply or marginal cost). Total surplus provides a measure of social efficiency: maximizing TS means achieving the greatest benefits possible given costs in a particular market.

Total surplus is maximized where demand equals supply at p^* and Q^* in Fig. 5.12. Notice that total surplus increases by area DFC when production increases from Q' to Q^*. Producing and consuming one more unit of output beyond Q^* will lower total surplus, however, due to the fact that the added benefit to consumers (reflected by the demand price) is lower than the added cost of producing it (reflected by the supply price, which equals marginal cost). Thus, total surplus is maximized where D = S or where the equilibrium price equals marginal cost. This description is exactly what is meant by allocative efficiency.

The discussion above makes it clear that perfect competition is allocatively efficient. A profit-maximizing firm in a competitive market equates price with marginal cost. In addition, demand equals supply in a competitive equilibrium. If all markets in the economy were perfectly competitive and there were no other market imperfections, then resources would automatically be allocated in an efficient manner without any need for government involvement. This outcome is the basis of Adam Smith's invisible hand theorem that he discussed in *The Wealth of Nations* (1776),

> [An individual] intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.

To summarize, a perfectly competitive market is statically efficient, all firms are technically and economically efficient, and the industry is productively and allocatively efficient.

---

[18] When PS is positive, as it is in this example, it is called economic rent. Even when economic profit is zero, firms may earn "economic rent." That is, an entrepreneur may enjoy owning a company so much that he or she is willing to earn only 8% on invested financial capital, even though the normal rate is 10%. If the owner does earn 10%, the 2% above the owner's opportunity cost is rent, not profit. Note that the concept of economic rent is different from the rent you pay on an apartment.

Before discussing market failure, we want to mention that Baumol et al. (1982) have proposed that a "perfectly contestable market" replace perfect competition as our benchmark of social efficiency. Three main assumptions drive the model of perfect contestability: (1) the sunk cost of entry is zero (i.e., there are no barriers to entry or exit), (2) entry is instantaneous, and (3) incumbent firms are slow to adjust price. Under these conditions, the theory implies that the threat of hit-and-run entry of a potential entrant will eliminate economic profits, even in a monopoly market. That is, the equilibrium price will equal average cost (but will be above marginal cost if there are economies of scale). This occurs because a rival enters whenever price exceeds average cost, with the entrant replacing the incumbent firm.

Unfortunately, the model of perfect contestability is deficient on two accounts. First, the assumptions of the model are false. In real markets, incumbents adjust price quickly and entry is slow. Moreover, Stiglitz (1987) showed that even small sunk costs overturn the zero profit result. Thus, hit-and-run entry is implausible. Second, the empirical evidence does not support the implications of the model, as we will discuss in later chapters. We summarize the perfectly contestable market model here because we will see in Chap. 20 that it provided one motivating factor in the deregulation movement of the 1970s and 1980s.

5.6 Market Failure and the Limitations of Perfect Competition

As we saw in the beginning of the chapter, a number of assumptions must hold for a market to be perfectly competitive and, therefore, efficient. When these conditions are not met, free markets generally fail to maximize total surplus. This is called **market failure**. There are four main sources of market failure: monopoly or market power, externalities, public goods, and a lack of information.[19]

Market power is associated with monopoly and imperfectly competitive markets that have too few competitors to guarantee competitive pricing. Specifically, a firm has market power if it has the ability to profitably maintain price above long-run marginal cost. When this happens, too little output is produced from society's perspective and allocative inefficiency results. The subject of market failure due to imperfect competition is the primary focus of this book, although we will refer to the other sources of market failure from time to time. In the next chapter, we discuss models in which strategic interaction is unimportant, the models of monopoly and monopolistic competition. The remainder of the book is devoted primarily to the study of competition in oligopoly markets that have more than one firm but fewer than many firms.

Another type of market failure occurs when goods are public rather than private. Pure **public goods** have two characteristics: nonrivalry in consumption

[19] As discussed in Chap. 1, just because markets fail to reach an ideal does not mean that government intervention can improve welfare. We ignore this issue here, but take it up more thoroughly in later chapters.

(one person's use does not diminish the quality or quantity of a good that is available to others) and nonexclusion (the cost of excluding people from using the good is uneconomic). Classic examples include broadcast radio programming, a lighthouse, and national defense. If Desmond listens to a broadcast radio program, you can also. In contrast, a loaf of bread is a private good; if Desmond eats a slice of bread, there are fewer slices available for you. Free markets will undersupply public goods. A private firm that tried to produce a public good would have difficulty earning a profit because (1) it is uneconomic to exclude users once the good is produced and (2) asking people who benefit to pay a share of the cost of producing the good will lead to underreporting given that consumers can receive the benefits whether they pay for it or not. Using a public good without paying for it is called the **free rider problem**. The presence of free riders leads to inadequate funding for public goods, which causes free markets to produce less than the socially optimal quantity of a public good and provides an important motive for government intervention.

An **externality** or spillover occurs when the action of one economic agent (producer or consumer) affects another without compensation. In the case of an external benefit or positive externality, such as a neighbor landscaping his or her front yard, you receive a benefit without having to pay for it. Similar to the public good situation, social benefits exceed private benefits to the neighbor and the act of landscaping will be undersupplied relative to the social optimum. Among producers, externalities are generally negative. The classic example is when a firm pollutes a river, which lowers the quality of drinking water to downstream consumers. This is called an external cost of production. In this case, the marginal cost of production to the firm is less than the marginal cost of production to society, which includes the cost to both the firm and to downstream consumers. Because a profit-maximizing firm will ignore the external cost to others, it will produce too much output and too much pollution from society's perspective. Thus, negative externalities will likely arise in a free market economy, providing a motive for government to impose taxes and regulations to mitigate the effects of externalities.

Externalities relevant to industrial organization can also be associated with consumers. One example is a **network externality**, which arises when one person's demand for a good depends on the number of others who consume the good.[20] Examples of goods and services where positive network externalities are present include cell phone plans, online dating services, and word processing software. If I have a document in one word processing software but you use a different software program, sharing files becomes more costly. By adopting the software package that you use, I benefit you as well as myself. Thus, my decision imposes a positive externality on you. A positive network externality such as this can be thought of as a bandwagon effect, which is discussed in Chap. 2. With a bandwagon effect, one person's demand for the good goes up as more and more consumers purchase it.

[20] For a review of the economics of networks, see Shy (2011).

A negative network externality is called a snob effect (see Chap. 2). In this case, the consumer who is a "snob" gets utility from being different from the crowd. Displeasure associated with the snob effect rises as a product becomes more popular. Thus, as the network grows in size, the value of the product declines for this consumer. From a social welfare perspective, network externalities tend to cause free markets to fail to reach the social optimum.

Market failure can also result when agents have **imperfect information** and a limited capacity to make choices in complex situations. If consumers do not know the prices of all suppliers, for example, higher-priced firms will not be driven out of the market and the socially optimal output level will not be produced. If consumers cannot distinguish high from low quality products before purchase, low quality goods (i.e., lemons) may drive high quality goods out of the market. Akerlof (1970) used this idea to explain why there are so few goods used cars available for purchase, which is why it is known as Akerlof's **lemons principle**.[21] Fortunately, the Internet has improved information flows dramatically, helping us to avoid some of these information shortfalls.

Cognitive weaknesses can cause people to fail to accurately assess information and can also lead to nonoptimal outcomes. Producers may manipulate information to their advantage when consumers are subject to framing, anchoring, and default effects, as discussed in Chap. 4. Consider a consumer who is working for a firm that offers three health insurance plans, A, B, or C. Suppose that an insurance company has convinced the firm to offer employees plan A as the default, even though plan B is best for most people. That is, the firm will provide plan A unless the employee overrides the default and chooses B or C. Studies show that people tend to choose the default even though it may not be in their best interest to do so. This creates a social problem because there is oversubscription to the default plan relative to plans B and C. Sellers also face difficulties when information is imperfect, such as identifying high and low quality job applicants or assessing the probability of success upon entering a market. Behavioral factors can also create problems for firms when their managers suffer from **overconfidence** or **over-exuberance**. We will see later in the book that free markets may fail to produce the social optimum when consumer or producer cognitive errors affect economic decisions.

5.7 Summary

1. The **assumptions of perfect competition** are that: firms produce perfectly homogeneous goods; there are many identical, profit-maximizing firms; there are no barriers to entry or exit; and there are no frictions or other forms of market imperfections.
2. The **firm demand** function (d) is horizontal at the market equilibrium price (p^*). Market demand and supply determine p^*.

[21] We will see that producers of high quality goods can solve this information problem by offering guarantees or warranties.

3. In perfect competition, the total revenue (TR) function is a ray from the origin with slope $= p$.
4. In perfect competition, MR = AR $= p^* = d$, where MR is marginal revenue and AR is average revenue of the firm.
5. The **profit equation** specifies the relationship between profit and output. The graph of the profit equation is concave or hill-shaped.
6. The output level that maximizes profit corresponds to the peak of the profit equation where the slope of a tangent line to the profit equation is zero. In the long run, this occurs where MR = MC. Since MR $= p$, we can also say that profit maximization occurs where $p =$ MC in perfect competition.
7. The firm's long-run **participation constraint** is that profit is nonnegative. If price is less than average cost, the firm is better off by producing nothing at all and exiting the industry.
8. When a perfectly competitive market is in long-run equilibrium, D $= S$ and p^* = MC = AC (i.e., firms earn zero economic profit). The market is in long-run disequilibrium when (economic) profits do not equal zero. When profits are positive, new firms will enter, increasing industry supply and reducing price. Firms will continue to enter until profits are zero (p^* = AC). When profits are negative, firms will exit the industry, decreasing supply and raising price. Firms will continue to exit until profits are zero (p^* = AC).
9. In a **constant-cost industry**, an increase in industry production does not affect input prices, and long-run industry supply is horizontal. In an **increasing-cost industry**, an increase in industry production pushes input prices up, and long-run industry supply has a positive slope.
10. The (ϵ_s) is the percentage change in quantity supplied resulting from a small percentage change in price, i.e., $\epsilon_s \equiv (\partial Q_s/\partial p)\ (p/Q_s)$.
11. **Comparative statics** is the analysis of the change in the equilibrium value of an endogenous variable of a model that results from a change in an exogenous variable or parameter. It involves a comparison of two "static" equilibria, ignoring the process that gets us from one equilibrium to another.
12. **Static efficiency** is a requirement of social welfare maximization, along with dynamic efficiency, equity, and macroeconomic stability. There are four types of static efficiency, two at the firm level (technical and economic efficiency) and two at the industry level (productive efficiency and allocative efficiency).
    - A firm **is technically efficient** when it uses the minimum quantity of inputs to produce a given output. A firm is **economically efficient** when it produces a given output at minimum cost.
    - An industry is **productively efficient** when it produces a given level of output at minimum cost. An industry is **allocatively efficient** when it produces the socially desirable quantity of output: the level where the marginal benefit to society, as reflected in the price, equals the marginal cost of production.
13. Perfect competition achieves all four types of static efficiency. Profit maximization guarantees economic efficiency (which requires technical efficiency).

Production efficiency occurs because every firm produces at minimum average cost. Allocative efficiency results because price equals marginal cost for all firms.

14. Social efficiency can be measured by **total surplus** which is the sum of consumer surplus and producer surplus. **Consumer surplus** is the difference between what a consumer is willing to pay for a good and the amount he or she actually pays. Graphically, consumer surplus is the area under the demand curve and above the price line. **Producer surplus** is the difference between the price a producer receives for an amount of a good and the marginal cost of production of the good. Graphically, producer surplus is the difference between the price line and the marginal cost curve.
15. Total surplus is the area under the demand curve and above the supply curve. Allocative efficiency is reached when total surplus is maximized. This occurs in perfect competition, because demand equals supply and price equals marginal cost. Total surplus reaches a maximum when demand equals supply and when price equals marginal cost.
16. Because perfect completion is socially efficient, the model of perfect competition serves as a **benchmark** for evaluating the social welfare implications of assumption violations, including alternative market structures.
17. **Market failure** occurs when private markets fail to produce the socially optimal level of output. Price does not equal the true opportunity cost of the resources used. The primary sources of market failure are market power, public goods, externalities, and imperfect information.
18. **Market power** exists when a firm can profitably maintain price above marginal cost, which is more likely in markets with few competitors. Quantity falls short of the social optimum and allocative inefficiency results.
19. Nonexclusion and nonrivalry in consumption are characteristics of **public goods**. Free riders can obtain use of the good without paying for it, and the public good is undersupplied.
20. **Externalities** are present when the actions of one agent affect others without compensation. Goods generating positive externalities are underproduced, while goods generating negative externalities are overproduced from society's perspective.
21. **Network externalities** occur when a person's demand for a good depends on the number of people who are using it. Bandwagon and snob effects are examples of network externalities.
22. **Lack of accurate information** about prices and product quality can lead to nonoptimal prices and output levels. Cognitive weaknesses can cause individuals to fail to accurately assess information and can also lead to nonoptimal outcomes. When there are framing, anchoring, and default effects, the demand function will not accurately represent preferences and the socially optimal output levels will not be produced. Overconfidence and over-exuberance can lead to entrepreneurial or managerial errors.
23. If consumers cannot distinguish high from low quality products before purchase, low quality goods may drive high quality goods out of the market. This is called Akerlof's (1970) **lemons principle**.

5.8 Review Questions

1. When a perfectly competitive firm is in long run equilibrium, will its average revenue equal minimum long run average cost? Explain.
2. Consider a good produced in a perfectly competitive market.

   A. Graph the industry supply and market demand curves, where the equilibrium price is \$4. Label the axes, curves, price, and quantity.
   B. For the firm, graph the corresponding demand function.
   C. Graph the corresponding total revenue function for the firm and indicate the slope and intercept on the graph.
   D. Graph the firm's average revenue and marginal revenue functions.

3. Consider a perfectly competitive market for a good in a constant-cost industry with market demand function given by $Q_D = 2,500 - 100p$, where Q_D is industry quantity demanded and p is market price. Producers have U-shaped AC curves that reach a minimum average cost at \$1 when 100 units of output are produced. In the long run, what will be the equilibrium price, industry level of production, industry profits, and number of firms?
4. State whether each statement is true or false and explain your answer.

   A. "Constant returns to scale" is another term for constant-cost industry.
   B. If economic profits are zero, firms will exit the industry.
   C. If a firm is minimizing costs, it must be maximizing profit.

5. Show graphically the MC, AC, MR, and q^* for a firm that does not meet the long-run participation constraint.
6. Suppose that a firm in a perfectly competitive, constant-cost industry sells 10,000 units at \$10 each. The firm's accounting profits are 10% above what would be considered a normal rate of return. Is \$10 the long-run equilibrium price? Why or why not?
7. Explain the long-run industry adjustment to a decrease in demand for an increasing-cost industry.
8. In Fig. 5.12, we found that producing less than the competitive level of output yielded less consumer, producer, and total surplus than at the competitive output level, Q^*. What would happen to the level of consumer surplus, producer surplus, and total surplus if the industry produced more than Q^* at a price equal to p^*?
9. Explain why the assumptions of profit maximization, product homogeneity, and free entry are necessary to assure that long-run profit is zero.
10. What is the difference between economic efficiency and productive efficiency?
11. Give an example of a public good, an external benefit, an external cost, and a network externality.
12. Explain how the presence of cognitive weaknesses among consumers, such as susceptibility to environmental cues for tempting or addictive products, can lead to market failure and problems with estimating market failure.