CHAPTER 3

Competition

Thou shalt not covet; but tradition approves all forms of competition.
—Arthur Hugh Clough

Perfect competition provides a benchmark against which the behavior of other markets is judged. The chapter starts by examining perfect competition, even though its strong assumptions apply to only a few markets. Next, two useful tools of analysis, elasticities and residual demand curves, are discussed. Then the chapter shows that competition has desirable efficiency and welfare properties. These properties, however, depend crucially on the assumptions of free entry and exit and on no externalities (firms bear the full costs of their actions). The chapter discusses the adverse effects if these conditions do not hold. The chapter concludes with examples of industries that most economists would characterize as reasonably competitive.

In this chapter, we stress five key points:

1. Perfect competition has many desirable properties.
2. Free entry and exit is a crucial factor in determining whether a market is perfectly competitive and efficient.
3. One important measure of welfare is maximized under perfect competition.
4. The desirability of perfect competition is reduced in the presence of externalities such as pollution.
5. Even if some of the necessary conditions for perfect competition do not hold, markets can come close to achieving the desirable properties of perfect competition.

Perfect Competition

Even though perfect competition is rarely, if ever, encountered in the real world, we study the perfect competition model because it provides an ideal against which to compare other models and markets. In later chapters, we examine how actual markets deviate from perfectly competitive ones, and determine which markets are likely to have the greatest deviations. The desirable properties of a perfectly competitive economy explain why economists generally favor competition. That a market deviates from the perfectly competitive model does not necessarily mean that the performance of a market can be improved, however, as is discussed at length throughout the book.

Assumptions

We define **perfect competition** as a market outcome in which all firms produce a homogeneous, perfectly divisible output; producers and consumers have full information, incur no transaction costs, and are price takers; and there are no externalities. That is, the main assumptions of perfect competition are:

- *Homogeneous Perfectly Divisible Output.* All firms sell an identical product. Consumers view the products of various firms as the same and hence are indifferent between them.
- *Perfect Information.* Buyers and sellers have all relevant information about the market, including the price and quality of the product. Firms can produce and consumers can buy a small fraction of a unit of output. As a result, the amount of output demanded or supplied varies continuously with price. This technical assumption avoids problems caused by large discrete changes in either supply or demand in response to small price changes.
- *No Transaction Costs.* Neither buyers nor sellers incur costs or fees to participate in the market.
- *Price Taking.* Buyers and sellers cannot individually influence the price at which the product can be purchased or sold. Price is determined by the market, so each buyer and seller takes the price as given.
- *No Externalities.* Each firm bears the full costs of its production process. That is, the firm does not impose externalities—uncompensated costs—on others. For example, pollution produced by a firm is an externality because the firm does not recompense the victims.

Some economists also assume that a perfectly competitive market has a large number of buyers and sellers. If there are many similar firms, no one firm can charge a price above the market price without losing all its customers, so the firm views the price at which it can sell as beyond its control. Similarly, consumers cannot find a firm willing to sell below the market price, so consumers must view the market price as beyond their control. Moreover, even if there are relatively few firms in a market, no firm can raise its price above the market price without losing all its customers if another firm can quickly enter the market and underprice it. Thus, because we assume firms and consumers are price takers, we do not also assume either that there are a large number of firms, or free entry and exit.[1] Competitive markets typically have a large number of firms and consumers, but industries can have all the properties of perfect competition even though there are few firms in those industries.

The Behavior of a Single Firm

Let us first examine the incentives of a typical firm. Suppose a firm has the short-run cost curves in Figure 3.1 and faces a market price of p_0. How much should it produce? Indeed, should it produce anything at all?

Profit Maximization. The objective of any firm, including a competitive firm, is to maximize its profits (or, equivalently, minimize its losses). The competitive firm's profits, π, are

$$\pi = pq - C(q),$$

where p is price, q is output, and $C(q)$ is total cost. As a result of the price-taking assumption above, the firm can sell all it wants at price p. (For example, the firm is too small a part of the market to influence the market price). That is, the firm faces a horizontal demand curve at price p.

It is profitable for the firm to expand output as long as the extra revenue from selling an additional unit exceeds the extra cost of producing that unit. The extra revenue from selling an additional unit is *price,* and the extra cost is the *marginal cost* (*MC*). That is, the optimal (profit-maximizing) production rule for a competitive firm is to expand its output until its marginal cost, *MC,* equals price, p.

Figure 3.1 illustrates the profit-maximizing decisions of a competitive firm facing a market price p_0. If the firm were producing a quantity greater than q_0, then p_0 would be less than *MC,* and the firm could increase its profits by reducing its output. If the firm were producing less than q_0, then p_0 would be greater than *MC,* and the firm could increase its profits by expanding its output. At output q_0, p equals *MC,* and profits are maximized.[2] In Figure 3.1, the shaded box represents profits.[3]

---

[1]We could derive the result that firms are price takers from these other assumptions. We make price taking an assumption for simplicity of presentation.

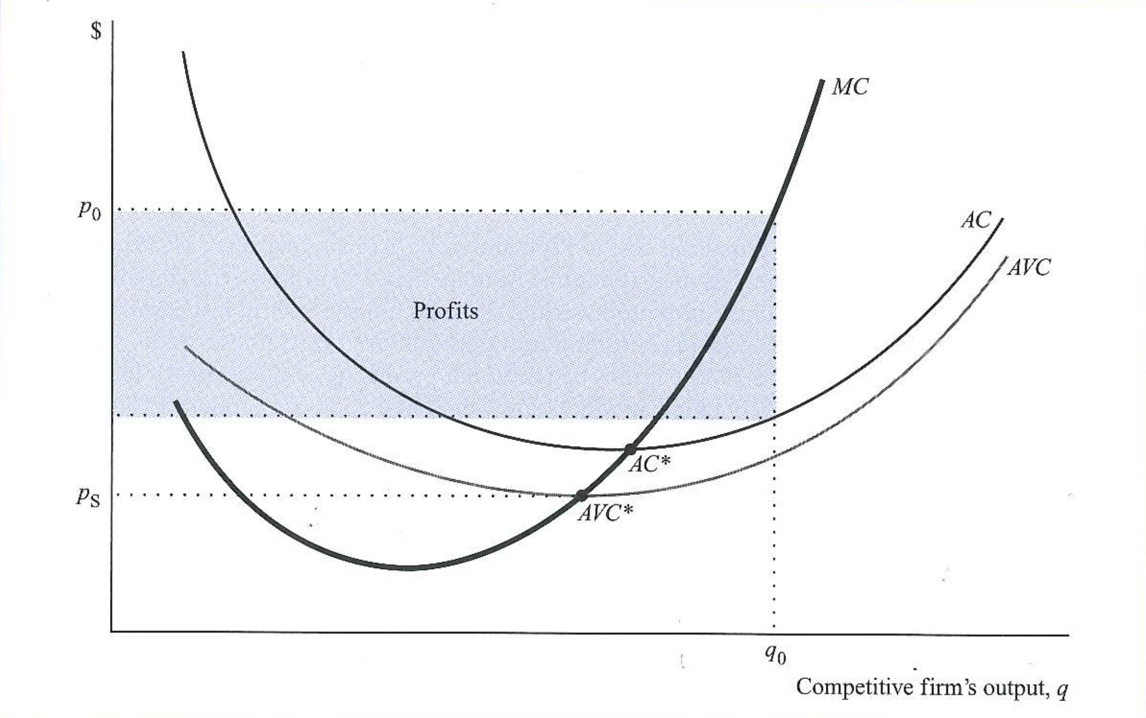
[2]The firm's objective is to

$$\max_q \pi = pq - C(q).$$

Its first-order condition is found by differentiating π with respect to q and setting that equal to zero: $p - C'(q) = 0$, where $C'(q) = dC(q)/dq$ is *MC*. This first-order condition—price equals marginal cost—is a necessary condition for profit maximization. The second-order condition is $C''(q) < 0$. That is, the second-order condition, which is a sufficient condition for profit maximization, is that the *MC* be upward sloping at the equilibrium.

[3]A firm's profits are total revenue minus total cost: $\pi = pq - C$, where pq (price times quantity) is total revenue. We can rewrite profits as average profits per unit (average revenue, p, minus average cost, $AC = C/q$) times the number of units sold (q), or $\pi = (p - AC)q$. Thus, profits can be shown graphically as a box with a height equal to average profits per unit, $p - AC$, and a length equal to the number of units, q, the firm sells.

FIGURE 3.1 Cost Curves and Profit Maximization



If the price, p, rises above p_0, the firm earns higher profits at its current output, but it earns even higher profits if it expands output until $p = MC$. If price falls below p_0, the firm earns lower profits at q_0, but it suffers less of a reduction if it reduces its output until $p = MC$. Thus, as the price rises, the firm moves up its marginal cost curve, and its profits rise; as the price falls, the firm moves down its marginal cost curve to minimize the reduction in its profits. Increases and decreases in profits signal a firm either to expand or contract output respectively.

Shutdown Decision. A firm produces only if doing so is more profitable than not producing. It produces only if the revenues from producing exceed **avoidable costs**: the costs that are not incurred if a firm ceases production. The revenues earned in excess of avoidable cost are called **quasi-rents**, which are the payments above the minimum amount necessary to keep a firm operating in the short run.

For simplicity, assume that all fixed costs are sunk. An example of a sunk cost given in Chapter 2 is that a firm is not refunded its incorporation fee if it ceases operating. In this case, avoidable costs are the same as variable costs. Thus, the rule for deciding whether to remain in business is: Produce and sell only if revenues are at least as great as total variable cost. Equivalently, the firm should produce and sell at price p only if p equals or exceeds average variable cost (AVC).

Minimum average cost (the lowest point on the *AC* curve), AC^*, is greater than minimum *AVC*, AVC^*, in the short run, because average costs are average variable costs plus average fixed costs. Thus, a firm finds it more profitable to produce than to shut down if price is below minimum average cost, $p < AC^*$, but above minimum average variable cost, $p > AVC^*$. It is more profitable to produce and earn some revenue in excess of variable cost than to shut down and earn no revenues (which can help offset the fixed costs). That is, the firm chooses to produce even though it is losing money when all costs are considered. Consider an example to clarify this apparent contradiction.

Suppose a firm's fixed cost is \$200 and sunk. Its marginal cost (*MC*) is constant at \$10 at quantities less than 100 units. At more than 100 units, *MC* is extremely high. If the price is \$10, the firm produces and sells 100 units. The firm just covers its production cost and makes no contribution to the \$200 fixed cost: It loses \$200.

If the price is \$9, the firm is better off not producing at all, because it loses an additional \$1 for every unit it produces and would lose \$300 if it produced 100 units. It is better to shut down and lose only \$200 than to produce and suffer greater losses.

If the price is \$11, by producing 100 units, the firm now more than covers variable cost: It earns \$100 above variable cost. It still loses money overall ($-200 + 100 = -100$) because of the fixed cost of \$200, but it is better to lose \$100 than \$200. The point of the example is that the decision to produce or not is *independent* of the fixed sunk cost. If fixed costs are sunk (incurred whether the firm produces or not), they should be ignored in deciding whether to produce.

If all fixed costs are sunk, a firm operates if p is greater than or equal to AVC^* but not if p is less than AVC^*. The price at which a firm ceases production is the **shutdown point**, which is p_s in Figure 3.1. That is, if price exceeds AVC^*, the firm operates along its *MC* curve. The **firm's supply curve** reflects the quantity that a firm is willing to supply at any given price. The competitive firm's supply curve, then, is the portion of the *MC* curve above AVC^*, the shutdown point.

If a firm suffers losses in the short run (the period in which costs are sunk), should it continue to operate and remain unprofitable in the future?[4] No. In the long run, a firm that is losing money will not reinvest—it will not continue to sink costs. Short-run losses are a signal that the firm should not invest further to replace plant and equipment. In the long run, a rational firm shuts down if it expects to have losses in each period forever. It prefers to cease production rather than invest in new facilities or maintenance and lose even more.

When a firm loses in the short run, its revenues are below the long-run opportunity cost of its resources. Because opportunity cost includes a normal profit, a firm that is making a loss may not literally be paying out more money than it receives; it is simply earning less than it could have earned had it invested its (already) sunk costs elsewhere.

[4]As described in Chapter 2, the short run and the long run are useful abstractions, but in reality adjustment costs determine how fast an industry can adjust to change. The time needed to adjust to any change depends on the current state of the industry and the size of the needed adjustment. See **www.aw-bc.com/carlton_perloff** "Adjustment Costs."

If fixed costs are not sunk, the shutdown decision depends on whether revenues exceed *avoidable* costs. An example (in Chapter 2) of an avoidable cost is the lawyer who can pay a penalty to break a lease. If some fixed costs are avoidable, a price equal to AVC^* is not high enough to prevent the firm from shutting down. Use the numbers from above and suppose that the fixed cost of $200 represents a yearly rental payment and that, for a $100 penalty fee, the landlord will release the lawyer from the obligation to pay $200. The firm compares losing $100 for sure (the penalty fee) with producing and earning revenues minus production costs minus the $200 rental payment. If price is $10, the firm earns $0 per sale and is stuck paying the $200 of fixed cost; therefore, it prefers to pay the $100 penalty and go out of business. Even if price were $10.50 so that the firm would make 50¢ on each of its 100 units sold, it would still be better to pay the $100 penalty and go out of business.

The price at which shutdown occurs is above average variable cost and closer to average cost the greater the proportion of fixed costs that are avoidable. In the extreme, when there are no sunk costs (all fixed costs are avoidable), the shutdown point coincides with the minimum point on the *AC* curve. Thus, if it has no sunk costs, a firm shuts down before it incurs economic losses.

The Competitive Market

Given the behavior of individual competitive firms, we can derive a market supply curve. The intersection of the market supply curve and the market demand curve determines the competitive equilibrium.

The Short-Run Equilibrium. We start by supposing that there are n identical firms and that all fixed costs are sunk in the short run. The short-run **market supply curve,** S in Figure 3.2b, is the horizontal sum of the supply curves of each firm, the *MC* curve above the minimum of the *AVC* curve in Figure 3.2a. The horizontal portion of the market supply curve reflects (1) that no output is forthcoming if price is below the shutdown point and (2) that at a price slightly above the shutdown point, all firms produce.

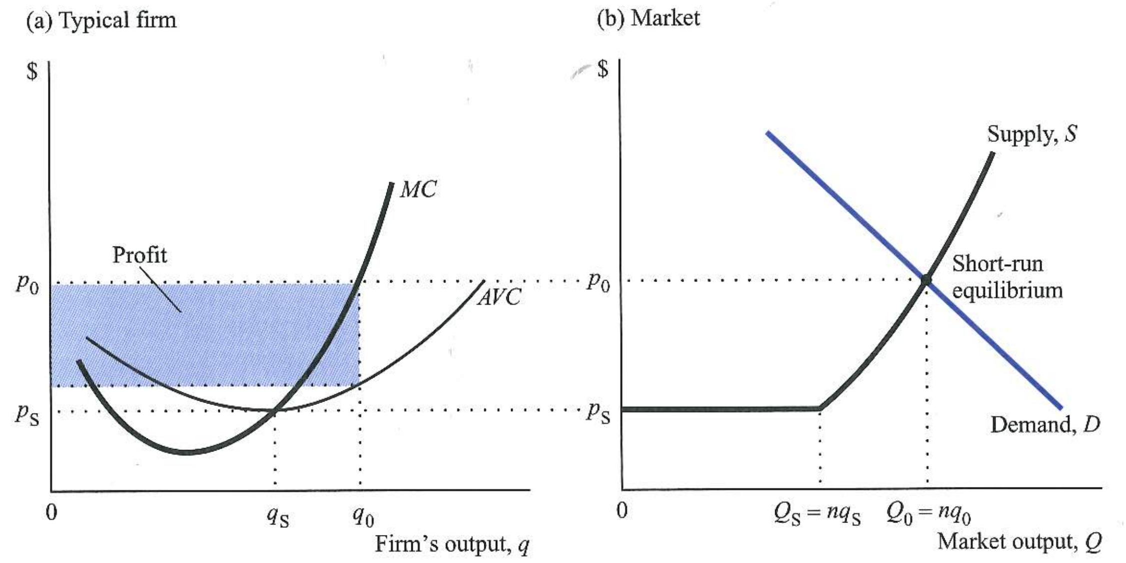
The intersection of the demand curve with the short-run market supply curve determines the *short-run equilibrium price,* p_0, and quantity, Q_0. The amount that firms want to supply at the equilibrium price exactly equals the amount that consumers demand at that price. There are no unsatisfied buyers and no unsatisfied sellers. All buyers pay and all sellers receive the same price.

In the short-run equilibrium in Figure 3.2, a typical firm may earn a profit, which provides an incentive for firms to enter the market. However, such entry cannot occur in the short run because firms cannot build new plants in the short run.

The Long-Run Equilibrium. In the long run, firms can adjust their levels of capital so that they can enter this market. Short-run profits or losses induce firms to enter or leave the market until price is driven to the minimum long-run average cost, AC^*, in the long run.

In Figure 3.2, firms are making a positive profit at the short-run equilibrium price p_0, which is determined by the intersection of the market demand curve and the original short-run market supply curve. In the long run, these profits induce new firms to

FIGURE 3.2 Short-Run Equilibrium



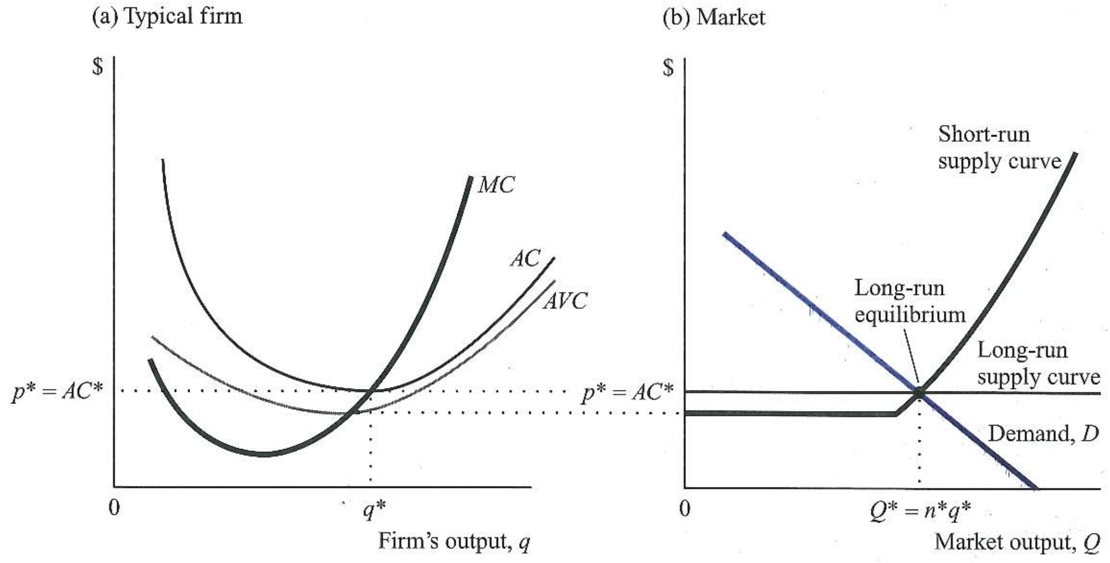
enter this market. If the number of firms that can potentially produce at the same cost is very large, the long-run supply curve is horizontal at the minimum of the average cost curve, AC^*, as Figure 3.3 shows. The long-run equilibrium is determined by the intersection of the demand curve and the long-run market supply curve. In Figure 3.3, the market is in a new short-run *and* long-run equilibrium because the demand curve, D, intersects both the long-run supply curve and the new short-run supply curve corresponding to the equilibrium number of firms, n^*. The equilibrium price is $p^* = AC^*$, and equilibrium output is $Q^* = n^*q^*$. In this long-run equilibrium, firms make zero profit.

Similarly, short-run losses induce firms to leave the market and reduce output until price rises again to yield normal (zero) profits. In long-run equilibrium, firms receive economic profits of zero, which is just enough to induce them to remain in the market.

The Slope of the Long-Run Supply Curve. In this last example, a very large number of firms could enter the market and produce at the same marginal and average costs as the existing firms. Consequently, the long-run perfectly competitive supply curve was perfectly flat at AC^*, which is the minimum average cost of production. However, the long-run supply curve need not be flat.

If an expansion of output causes the prices of some key inputs to rise, the long-run supply curve tends to be upward sloping. As the output of wheat produced increases, farmland becomes more valuable, and the land rents (or the opportunity cost of owning the land) increase. As rents increase, the average cost curve of each farmer rises so the minimum average cost, AC^*, increases. Thus, the long-run supply curve for the

FIGURE 3.3 Long-Run Equilibrium



wheat market (whose height is traced out by the minimum average cost points) rises as output expands.

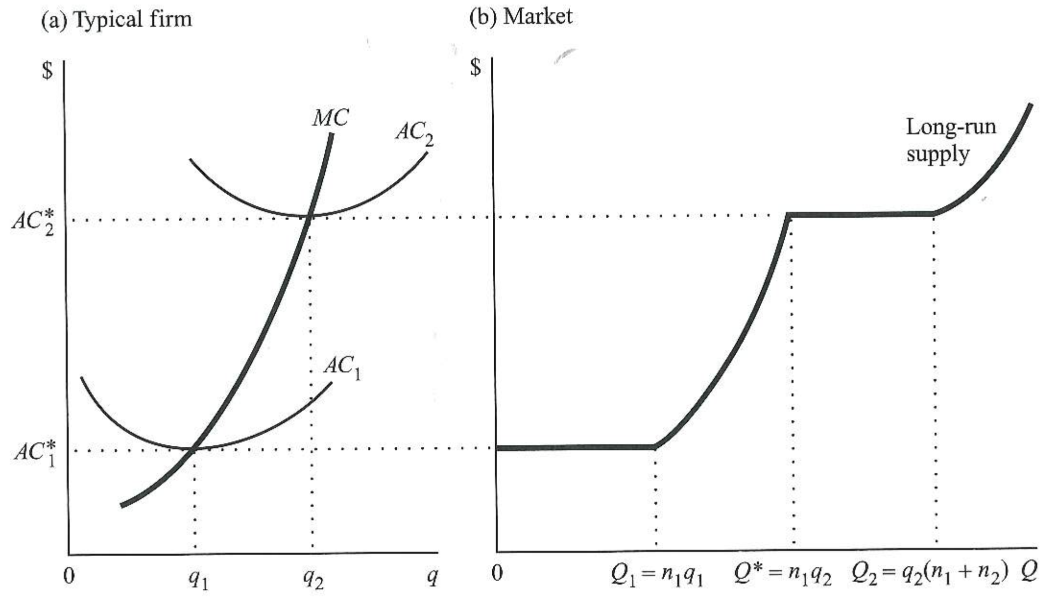
Whenever some factors of production (such as fertile land) are in fixed supply, their price gets bid up as market output expands. Prices of key inputs may fall as output expands if there are economies of scale. If input prices fall as output expands, the long-run market supply curve could slope down. The long-run supply curve of a market tends to be flat as long as the market accounts for only a small fraction of any one factor's total employment.

Another reason why the long-run supply curve may be upward sloping is that there are only a few firms that can produce at low costs. For market output to increase, less efficient firms have to enter the market. In Figure 3.4a, there are n_1 low-cost, efficient firms with marginal cost curve MC and average cost curve AC_1. The minimum point on AC_1 is AC^*_1, which is obtained if the firm produces q_1 units of output. For market output levels up to $Q_1 = n_1 q_1$, these low-cost firms can produce at the minimum average cost, AC^*_1, so the long-run supply curve is flat at AC^*_1 up to Q_1, as Figure 3.4b shows. If less than Q_1 is demanded, some of these n_1 firms exit the market.

If the market demand is slightly larger than Q_1, the average cost of production must rise. The market supply curve is the horizontal sum of the supply curves of the n_1 firms: Their marginal cost curves above AC^*_1. Thus, because there are no more low-cost firms, the market supply curve rises beyond Q_1.

Now suppose that there are n_2 other firms that can produce this product with the same marginal cost curve as the first n_1 firms but with an average cost curve, AC_2, with a higher minimum average cost, AC^*_2 ($> AC^*_1$). That is, these high-cost firms have larger fixed costs than do the low-cost firms.

FIGURE 3.4 Upward-Sloping Long-Run Supply Curve



If the quantity demanded is slightly greater than $Q^* = n_1q_2$, the price is AC^*_2 and some high-cost firms enter the market. Increases in market demand beyond this point are met by additional high-cost firms entering the market and producing q_2 at an average cost of AC^*_2. When the quantity demanded can no longer be met by entry by additional high-cost firms, the long-run supply curve again rises, tracing out the sum of the marginal cost curves of all the firms in the market. That is, the long-run supply curve rises for output greater than $Q_2 = Q^* + n_2q_2 = q_2(n_1 + n_2)$.

If the quantity demanded exceeds n_1q_1 but is less than Q^* (that is, the second group of firms has not entered the market), the low-cost firms earn an unusual return (profit) to their scarce knowledge or other scarce resource that enables them to produce at relatively low costs. That is, they earn a **rent**: a payment to the owner of an input beyond the minimum necessary to cause it to be used. If the quantity demanded exceeds Q_2, both types of firms earn rents on their scarce know-how or other scarce input.[5]

[5]In some markets where firms must incur substantial sunk costs so that only a few firms can efficiently produce, a competitive equilibrium may be impossible. In such markets where no competitive equilibrium exists, the market exhibits instability, including price wars and bankruptcies. Here, firms may temporarily make above-normal profits; these profits attract other firms, the additional competition results in all the firms in the market making losses, some firms exit, the remaining firms make above-normal profits, and the process repeats. The study of when interactions between firms and consumers will lead to stability is called the theory of the core. See Clark (1923), Telser (1978), and **www.aw-bc.com/carlton_perloff** "Nonexistence of Competitive Equilibrium."

Elasticities and the Residual Demand Curve

Throughout the rest of this book, we make repeated use of two related concepts to analyze both competitive and noncompetitive industries: (1) the price elasticity of demand or supply, and (2) the demand curve facing a single firm, which is called the firm's *residual demand curve.* The price elasticity of supply or demand aids in understanding how a market responds to changes in either demand or supply. The residual demand facing a single firm allows an analyst to comprehend the behavior of a single firm. We now examine how the elasticity of the residual demand curve is related to the assumption that a competitive firm cannot affect price.

Elasticities of Demand and Supply

If either the demand or the supply curve shifts, the competitive equilibrium changes, and the shapes of the demand and supply curves influence how the new equilibrium compares to the old. For example, if the demand curve is perfectly flat, the competitive price remains unchanged, even if the supply curve shifts radically.

One concept used to characterize the shape of demand or supply curves is the price elasticity of demand or supply (often the word *price* is omitted). The **elasticity of demand** is the percentage change in quantity demanded in response to a given small percentage change in price.[6] Similarly, the **elasticity of supply** is the percentage change in quantity supplied in response to a given small percentage change in the price. The elasticity of demand is always a negative number, and the elasticity of supply is usually, but not always, positive.

If a 1 percent increase in price leads to a more than a 1 percent reduction in the quantity demanded (so that the total amount paid in the market falls), a demand curve is called **elastic**. That is, an elastic demand curve has an absolute value of the elasticity of demand greater than 1 (the absolute values of 1 and -1 are both 1). It is common to omit the phrase *the absolute value of* when discussing the price elasticity of demand. The statement, "The price elasticity of demand is 2," is interpreted to mean that the price elasticity is -2.

When the absolute value of the elasticity of demand is 1, the demand curve is said to have **unitary elasticity**. In that case, a 1 percent change in price causes a 1 percent change in the quantity demanded, and the total amount paid (total revenues) remains constant. If the absolute value of the elasticity of demand is less than 1, the demand curve is **inelastic**: A 1 percent increase in price causes less than a 1 percent decline in the quantity demanded, and the total amount paid rises.

[6] The price elasticity of demand at price p and quantity Q is the percentage change in quantity divided by the percentage change in price (if that change is small): $(\Delta Q/Q)/(\Delta p/p) = (p/Q)(\Delta Q/\Delta p)$. Because the elasticity is the ratio of two percentage terms, the elasticity is invariant to changes in scale of either price or quantity (it is a *pure* number—without scale itself). For example, if price is measured in cents rather than in dollars, the elasticity is unchanged even though the slope of the demand curve, $\Delta Q/\Delta p$, does change. The technical definition of the price elasticity is $(p/Q)(dQ/dp)$.

In general, the elasticities of demand and supply depend upon many economic factors, such as the level of output, the availability of substitute products, and the ease with which suppliers can alter production. For example, as more substitute products are available, consumers find it easier to substitute for a product if its price rises, which makes its demand curve more elastic. Similarly, the more flexible the production process of a firm, the more likely it is that the firm can greatly increase production in response to a price increase, which tends to increase the elasticity of supply.

The Residual Demand Curve of Price Takers

Competitive firms are often described as *price takers.* They believe that they cannot affect the market price and must accept, or take, it as given. There are three equivalent ways to describe a firm's inability to affect price, all of which are used in this chapter:

- A competitive firm is a price taker.
- The demand curve facing a competitive firm is horizontal at the market price.
- The elasticity of demand facing a competitive firm is infinite.

A firm is a price taker if it faces a horizontal demand curve, because a horizontal demand curve has an infinite price elasticity of demand. If a firm facing an infinite price elasticity raises its price even slightly, it loses all its sales. Equivalently, by lowering its quantity, the firm cannot cause the price to rise. In contrast, a firm facing a downward-sloping demand curve can raise its price by decreasing its output.

If the number of firms in a market is large, the demand curve facing any one firm is nearly horizontal (elasticity of demand is infinite) even though the demand curve facing the market is downward sloping with a low elasticity. Indeed, for most market demand curves, there do not have to be very many firms in a market for the elasticity of demand facing a particular firm to be large.

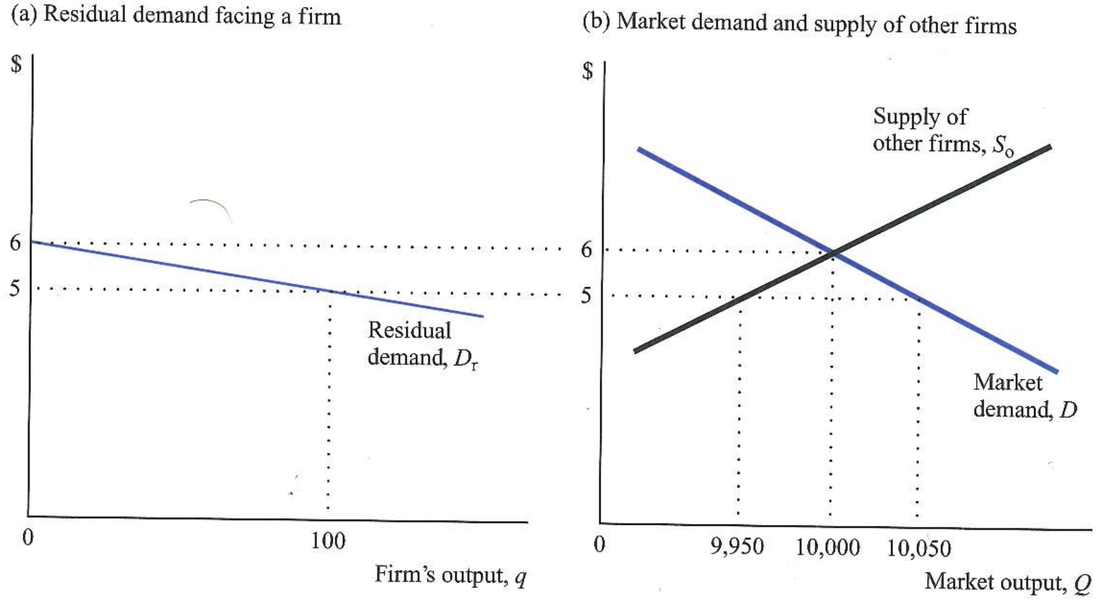
To show this result, it is necessary to determine the demand curve facing a particular firm: the **residual demand curve.** A firm sells to people whose demands are not met by the other firms in the market. For positive quantities of residual demand, the residual demand, $D_r(p)$, is the market demand, $D(p)$, minus the supply of other firms, $S_o(p)$:

$$D_r(p) = D(p) - S_o(p).$$

If $S_o(p)$ is greater than $D(p)$, $D_r(p)$ is zero.

Figure 3.5b shows the market demand curve and the supply curve of all the firms except one. Figure 3.5a shows the residual demand curve facing a particular firm, which is the horizontal difference between the quantity demanded by the market at a given price minus the supply of other firms at that price. For example, at a price of \$5, market demand is 10,050 units and the supply of the other firms is 9,950 units in Figure 3.5b. Thus, market demand exceeds their supply by 100 units at \$5, so the remaining firm faces a residual demand of 100 units at that price.

FIGURE 3.5 Derivation of Residual Demand Curve



At \$6, the supply of other firms equals the market demand. The residual demand facing the firm in Figure 3.5a is zero. Were the price to rise even higher, other firms would be willing to supply even more than is demanded. Thus, at any price greater than or equal to \$6, the firm in panel a sells no units.

The residual demand curve facing the firm, Figure 3.5a, is much flatter than the market demand curve in Figure 3.5b. Similarly, the single firm's demand elasticity is much higher than the market elasticity. For example, the elasticity of demand for the individual firm is −11 at \$5.50, whereas the corresponding market elasticity of demand is approximately −0.027.[7] In other words, the firm's residual demand curve is over 400 times as elastic as the market demand curve at this price.

More generally, if there are n identical firms in the market, then the elasticity of demand facing Firm i is

$$\epsilon_i = \epsilon n - \eta_o(n - 1), \tag{3.1}$$

[7]For a linear demand curve, $Q = a - bp$, the elasticity of demand is the slope of the demand curve, $dQ/dp = -b$, times p/Q. Thus, the elasticity of demand of the residual demand curve is $-100(5.50/50) = -11$ and the elasticity of demand of the market demand curve is $-50(5.50/10{,}025) \approx -0.027$.

where ϵ is the market elasticity of demand (a negative number), η_o is the elasticity of supply of the other firms (a positive number), and $(n - 1)$ is the number of other firms.[8]

Thus, for a given market elasticity, as the number of firms in a market, n, increases, the elasticity facing a single Firm i, ϵ_i, grows large in absolute value (more negative). Similarly, the larger the elasticity of supply of the other firms, η_o, or the more other firms there are, the larger in absolute value (more negative) is the elasticity of demand facing Firm i.

Table 3.1 shows how the elasticity of demand facing a single firm varies with the number of firms and the market elasticities, given that the supply of other firms is completely inelastic ($\eta_o = 0$). For example, if the market elasticity is unitary ($\epsilon = -1$) and there are 50 firms, then $\epsilon_i = -50$. That is, if a firm were to increase its price by 1 percent, the quantity it sells would fall by about 50 percent. If the market demand elasticity is -0.5 and there are 1,000 firms, $\epsilon_i = -500$, so that if the firm were to raise its price by a tenth of a percent, the quantity it sells would fall by 50 percent. Thus, even if the supply of the other firms is completely inelastic, if there are a large number

TABLE 3.1 **Price Elasticity for a Single Firm**

	Market Elasticity		
Number of Firms n	Inelastic $\epsilon = -0.5$	Unitary $\epsilon = -1$	Elastic $\epsilon = -5$
10	−5	−10	−50
25	−12.5	−25	−125
50	−25	−50	−250
100	−50	−100	−500
500	−250	−500	−2,500
1,000	−500	−1,000	−5,000

Note: Because the supply of the other identical firms is assumed to be perfectly inelastic ($\eta_o = 0$), the elasticity of demand facing a particular firm is $\epsilon_i = n\epsilon$.

[8]The residual demand curve facing any one firm is $D_r(p) = D(p) - S_o(p)$. Differentiating $D_r(p)$ with respect to p, we obtain

$$\frac{dD_r}{dp} = \frac{dD}{dp} - \frac{dS_o}{dp}.$$

Let the quantity produced by one firm be $q = Q/n$ and the total quantity produced by all the other firms be $Q_o = (n - 1)q$. Multiplying both sides of the expression above by p/q and multiplying and dividing the first term on the right-hand side by Q/Q and the second term by Q_o/Q_o, this expression is

$$\frac{dD_r}{dp}\frac{p}{q} = \frac{dD}{dp}\frac{p}{Q}\frac{Q}{q} - \frac{dS_o}{dp}\frac{p}{Q_o}\frac{Q_o}{q},$$

where $q = D_r(p)$, $Q = D(p)$, and $Q_o = S_o(p)$. This expression is rewritten as $\epsilon_i = \epsilon n - \eta_o(n - 1)$ in Equation 3.1.

EXAMPLE 3.1 *Are Farmers Price Takers?*

In most U.S. agricultural markets there are a large number of farms, and no farm has as much as even 1 percent of total sales. As a result, the elasticity of demand facing each farm is extremely large. Farms are price takers.

We can roughly calculate the residual demand price elasticity facing an individual farm. For simplicity, we assume that other farms have an inelastic supply ($\eta_o = 0$), which may be a reasonable assumption in the short run. Less accurately, we assume that all farms are approximately the same size, so that each farm's share of the market is equal to 1 divided by the number of farms. The following table shows the approximate elasticity of demand facing each farm.

Crop	Estimated Market Demand Elasticity	Number of Farms	Each Farm's Residual Demand Elasticity
Fruits			
apples	−.20	28,160	−5,620
grapes	−1.03	19,961	−20,560
peaches	−.82	14,459	−11,856
Vegetables			
asparagus	−.65	2,672	−11,140
cucumbers	−.30	6,821	−2,046
dry onions	−.16	3,296	−527
sweet peppers	−.25	6,271	−1,568
tomatoes	−.38	14,366	−5,459

Sources: Number of Farms: U.S. Department of Commerce, Bureau of the Census, 1997 Census of Agriculture; Survey of Elasticities: You; Epperson, and Huang (1998).

Thus, each farm faces a gigantic price elasticity. For example, were a grape farm to increase its price by as little as 0.001 percent (one-thousandth of one percent), the quantity demanded from the farm would fall by 21 percent. Each farm is a price taker.

of firms in the market, the elasticity of demand facing a single firm is very large, as Example 3.1 illustrates.

Efficiency and Welfare

The welfare of the people is the chief law. —*Cicero*

The competitive equilibrium has desirable efficiency and welfare properties. Indeed, no one can be made better off without making someone else worse off in the competitive equilibrium.

Efficiency

The competitive equilibrium of price and quantity has two desirable efficiency properties. First, production is efficient in the sense that there is no possible rearrangement of resources (such as labor, machines, and raw materials) among firms that can increase the output of one product without reducing the output of at least one other product.

Second, consumption is efficient. The value that a buyer places on consuming the good is exactly equal to the marginal cost of producing that good (remember, the competitive price equals the marginal cost of production). Moreover, no rearrangement of goods among consumers can benefit a consumer without harming at least one other consumer.

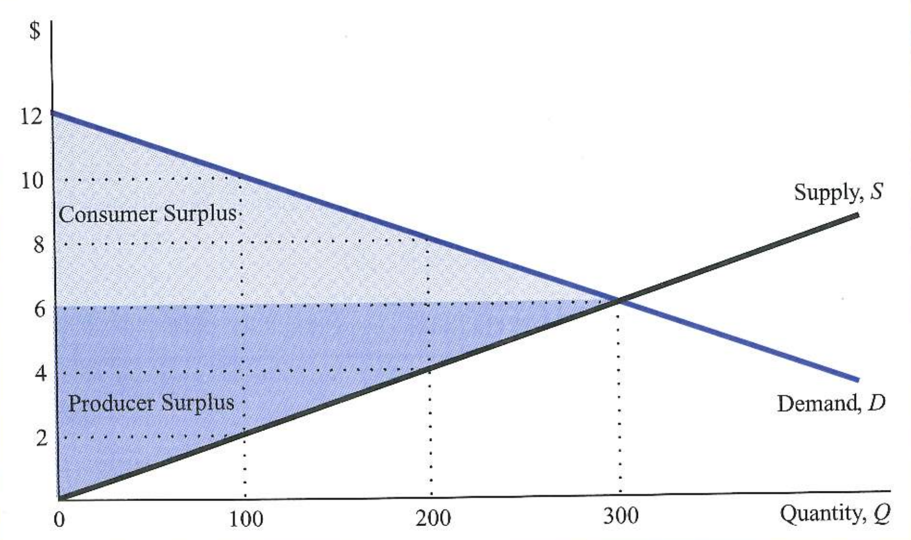
Welfare

We now describe a common measure of welfare, show that competition maximizes this measure of welfare for any given distribution of income, and illustrate that departures from competition lower welfare. In particular, in the next section, we show that restrictions that prevent firms from entering a market lower welfare.

Consumer Surplus. Typically, consumers value the goods they purchase above the amount they actually pay for them. **Consumer surplus** is the amount above the price paid that a consumer would willingly spend, if necessary, to consume the units purchased.

A good's demand curve reflects the value that consumers place on consuming additional units of a good. For example, the demand curve in Figure 3.6 indicates that

FIGURE 3.6 Consumer Surplus and Producer Surplus



consumers would pay \$10 for 100 units of the good, \$8 for 200 units, and \$6 for 300 units.

In the competitive equilibrium in Figure 3.6, consumers pay \$6 for 300 units. They would have been willing to spend \$4 more for the first 100 units, \$2 more for the first 200 units, and no extra amount for 300 units. The total consumer surplus is the shaded area below the demand curve and above the equilibrium price of \$6 up to the equilibrium quantity of 300 units.[9] This area equals \$900 (= [\$12 − \$6] × 300/2).

In the competitive market, consumers paid \$1,800 to buy the 300 units. In this example, consumer surplus is 50 percent of the amount they actually pay. If consumers could have had the choice of buying 300 units or none, they would have been willing to spend up to \$2,700 (the \$1,800 they spent, plus the extra \$900 in consumer surplus) to purchase the 300 units.

Producer Surplus. Similarly, firms may receive more for the goods they sell than it costs them to produce those goods. **Producer surplus** is the largest amount that could be subtracted from a supplier's revenues and yet the supplier would still willingly produce the product.

We can use information from the supply curve to calculate firms' producer surplus. A supply curve represents the marginal cost of producing output. For example, in Figure 3.6, it costs firms \$2 to produce 100 units, \$4 to produce 200 units, and \$6 to produce 300 units. The producer surplus is the area above the supply curve and below the market price up to the quantity sold. The producer surplus is the area equal to \$900, which is above the supply curve and below the price of \$6 up to 300 units. That is, firms would be willing to pay \$900 for the right to sell 300 units of the good at \$6 rather than selling none at all.

Welfare. One common measure of welfare from a market is the sum of consumer surplus and producer surplus. This measure of welfare is the value that consumers and producers would be willing to pay to purchase the equilibrium quantity of output at the equilibrium price.

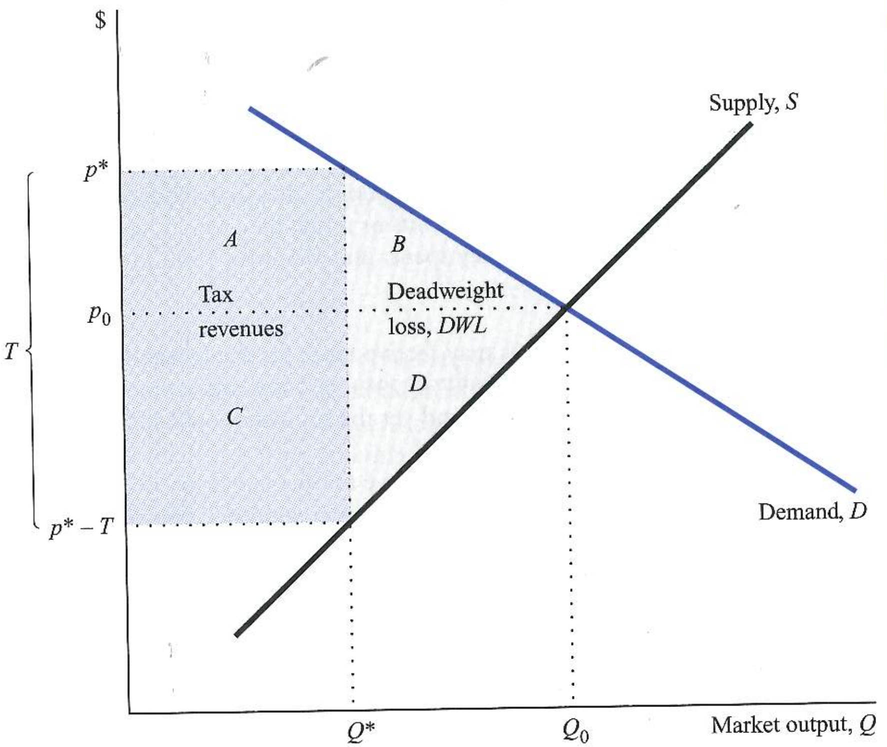
Figure 3.6 illustrates that this measure of welfare is maximized at the competitive equilibrium. For example, if fewer units were produced, welfare would fall, as we now show.

Deadweight Loss. The cost to society of a market's not operating efficiently is called **deadweight loss** (*DWL*). It is the welfare loss—the sum of the consumer surplus and producer surplus lost—from a deviation from the competitive equilibrium.

For example, the competitive equilibrium is at price p_0 and quantity Q_0 in Figure 3.7. At Q_0, the value that a consumer places on additional consumption equals the marginal cost of producing the good. If the government taxes this good or restricts its

[9]Consumer surplus is an accurate measure of consumer well-being if there are no income effects (a change in a consumer's income leaves demand unchanged). Even when there are income effects, changes in consumer surplus can provide a close approximation to changes in welfare (Willig 1976).

FIGURE 3.7 Deadweight Loss from a Tax



sale, this link between the value the consumer places on an additional unit and the cost of producing it is broken, which lowers welfare.

Suppose that the government collects a tax of T per unit of goods sold. If a customer pays p, the government takes T of that, and the firm receives $p - T$. Thus, the tax creates a wedge of T between the value that the marginal demander places on the good (as shown by the demand curve) and the cost that the marginal supplier is willing to incur to produce the good (as shown by the supply curve). The imposition of the tax reduces the quantity sold from Q_0 to Q^*. The price consumers pay rises to p^*, and the price firms receive falls to $p^* - T$.

In this after-tax equilibrium, the amount sold, Q^*, is less than in the competitive equilibrium, Q_0, and the value consumers place on consuming an additional unit, p^*, now exceeds the marginal cost of producing it by T. Consumers suffer a loss of consumer surplus equal to areas A and B (the area between p^* and p_0, to the left of the demand curve). Suppliers suffer a loss of producer surplus equal to areas C and D (the area between p_0 and $p^* - T$ to the left of the supply curve). The government receives tax revenues equal to TQ^*, boxes A and C. Thus, the transfer from consumers and producers to the government (the tax revenues = boxes A and C) is less than the combined loss of the consumers and producers. This extra cost to society due to reduced output is the deadweight loss, which equals the sum of the triangles B and D in Figure 3.7.[10]

The deadweight loss triangle is the total loss to society if the government makes good use of the tax revenues. The *DWL* triangle is an efficiency loss because the marginal cost of producing a good is less than the marginal willingness of consumers to pay for it.

As long as the government makes efficient use of this money, the tax revenues are not an efficiency loss. Rather, the tax revenues reflect a redistribution of income from buyers and sellers of this good to those who benefit from the government's use of these funds.

Entry and Exit

As we show throughout the rest of this book, ease of entry and exit plays a critical role in determining market structure and the subsequent performance of firms. If firms that are as efficient as those already in the market cannot easily enter, existing firms may be able to exercise market power by setting prices above marginal cost.

Restrictions on Entry

In many industries, governments or groups of firms collectively set licensing requirements that restrict entry (see Example 3.2). An example of an entry restriction is the limit on the number of taxicabs allowed in many cities throughout the world. Such entry restrictions elevate prices above competitive levels.

Figure 3.8 illustrates how a restriction on entry leads to a price above the long-run competitive equilibrium price. In this market, a large number of firms could produce with identical cost curves, as panel a shows.

Figure 3.8b shows two long-run supply curves for a market where all firms have identical costs. In the absence of a government restriction on entry, there are 150 firms in this market. The competitive equilibrium is determined by the intersection of the supply curve for 150 firms and the market demand curve. The equilibrium price is p_0, and each firm is producing at the minimum of its long-run average cost curve, AC^*.

[10]The deadweight loss triangle can be expressed in terms of the elasticities of demand and supply. For simplicity, assume that the supply curve is perfectly horizontal (infinitely elastic). Then the deadweight loss triangle $= -1/2\Delta p\Delta Q$, where $\Delta p = p^* - p_0$ and $\Delta Q = Q^* - Q_0$. The elasticity of demand, ϵ, is (approximately) equal to $(\Delta Q/Q_0)(p_0/\Delta p)$. Define t as $\Delta p/p_0$, which is the percentage change in the price due to the tax. The deadweight loss triangle equals

$$-\frac{1}{2}\Delta p\Delta Q \approx -\frac{1}{2}\frac{\Delta p}{p_0}p_0Q_0\left(\frac{\Delta Q}{Q_0}\frac{p_0}{\Delta p}\right)\frac{\Delta p}{p_0} \approx -\frac{1}{2}t^2R\epsilon,$$

where R is the revenue, p_0Q_0, and $\approx$ means *approximately equal*. Thus, the deadweight loss depends on the size of the market, R, in addition to t and ϵ.