3 BASIC PRICE OPTIMIZATION

In this section we introduce the basic elements of pricing and revenue optimization and show how the basic pricing revenue and optimization problem can be formulated as an optimization problem. The goal of the optimization problem—its objective function—is to maximize contribution: total revenue minus total incremental cost from sales. The key elements of this problem are the price-response function and the incremental cost of sales, both of which we introduce in this section. We then formulate and solve the pricing and revenue optimization function in the case of a single product in a single market without supply constraints and derive some important optimality conditions.

3.1 THE PRICE-RESPONSE FUNCTION

A fundamental input to any PRO analysis is the *price-response function*, or *price-response curve*, $d(p)$ which specifies how demand for a product varies as a function of its price, p. There is one price-response function associated with each element in the PRO cube —that is, there is a price-response function associated with each combination of product, market-segment, and channel. The price-response function is similar to the market demand function found in economic texts. However, there is a critical difference. The price-response function specifies *demand for the product of a single seller as a function of the price offered by that seller*. This contrasts with the concept of a *market demand curve*, which specifies how an entire market will respond to changing prices. The distinction is critical because different firms competing in the same market face different price-response functions. Referring to Table 1.2, the price-response function facing Amazon for *Bag of Bones* is likely to be quite different from the one facing ecampus.com. The differences in the price-response functions faced by different sellers are the result of many factors, such as the effectiveness of their marketing campaigns, perceived customer differences in quality, product differences, and location, among other factors.

In a perfectly competitive market, the price response faced by an individual seller is a vertical line at the market price, as shown in Figure 3.1. If the seller prices above the market price, his demand drops to 0. If he prices below the market price, his demand is equal to the entire market. A standard example used in economics texts is wheat:

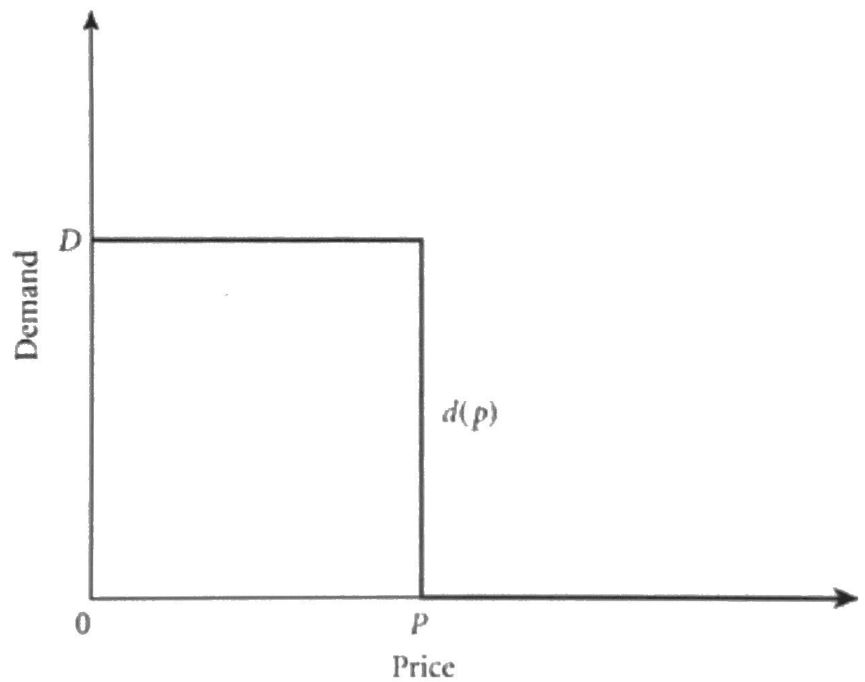

Figure 3.1 Price-response curve in a perfectly competitive market.

> The best example to keep in mind is that of a wheat farmer, who provides a minuscule percentage of the wheat grown in the world. Regardless of whether he produces 10 bushels or 1,000, he remains too small to have any impact on the going market price....If he tries to charge even a fraction of a penny more, he will sell no wheat, because buyers can just as easily buy from someone else. If he charges even a fraction of a penny less, the public will demand more wheat from him than he can possibly produce—effectively an infinite quantity.1

In other words, wheat is a commodity—buyers are totally indifferent among the offerings of different sellers, they have perfect knowledge about all prices being offered, and they will buy the product only from the lowest-price seller. Furthermore, each seller is small relative to the total size of the market. In this situation, the seller has no pricing decision—his price is set by the operation of the larger market. To quote a popular text: "In a competitive market, each firm only has to worry about how much output it wants to produce. Whatever it produces can only be sold at one price: the going market price."2 At any price below the market price, the demand seen by a seller would be equal to the entire demand in the market—the amount *D* in Figure 3.1. At any price above the market price he sells nothing. The seller of a true commodity in a perfectly competitive market has no need for pricing and revenue optimization—indeed, he has no need of any pricing capability whatsoever. However, true commodities are surprisingly rare. The vast majority of companies face finite customer responses to price changes and therefore have active PRO decisions.

The price-response functions that we consider—and those facing most companies most of the time—demonstrate some degree of smooth price response. An example is shown in Figure 3.2. Here, as price increases, demand declines until it reaches zero at some *satiating price P*. This type of smooth market-response function is usually termed a *monopolistic* or *monopoly* demand curve in the economics literature. The terminology is somewhat unfortunate, since the fact that a company faces some level of price response hardly means that it is a "monopoly." Companies such as United Airlines, UPS, and Ford all face smooth price-response functions for their products, yet none of them would be considered a monopoly in the usual sense of the world.

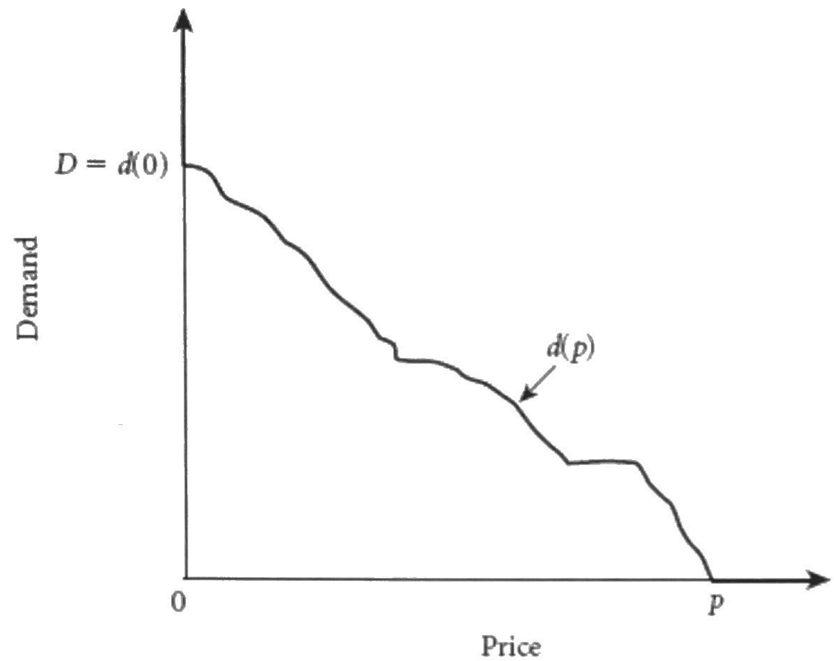

Figure 3.2 Typical price-response curve.

3.1.1 Properties of the Price-Response Function

The price-response functions used in PRO analysis have a time dimension associated with them. This is in keeping with the dynamic nature of PRO decisions—we are not setting prices that will last "in perpetuity" but prices that will be in place for some finite period of time. The period might be minutes or hours (as in the case of a fast-moving e-commerce market), days or weeks (as in retail markets), or longer (as in long-term contract pricing). At the end of the period we have the opportunity to change prices. The demand we expect to see at a given price will depend on the length of time the price will be in place. Thus, we can speak of the price-response function for a model copier over a week or over a month, but without an associated time interval there is no single price-response function.

There are many different ways in which product demand might change in response to changing prices and, thus, many different possible price-response functions. However, all of the price-response functions we consider will be nonnegative, continuous, differentiable, and downward sloping.

Nonnegative. We assume that all prices are greater than or equal to zero; that is, $p \geq 0$.[3]

Continuous. We assume that the price-response function is *continuous*—there are no “gaps” or “jumps” in the market response to our prices. More formally, if $d(0) = D$ and P is the satiating price, that is, the lowest price for which $d(P) = 0$, then for every $0\ q \leq D$ there is a price $p \leq P$ such that $d(p) = q$. This implies that there is a price that will generate any level of demand between 0 and the maximum demand. This property, known as *invertability*, is often very useful.[4]

Differentiable. Differentiability means that the price-response function is smooth, with a well-defined slope at every point.[5] As with the assumption of continuity, this assumption involves taking a mathematical liberty, since prices are only defined for fixed increments. However, differentiability allows us to use the tools of calculus to solve the constrained optimization problems that arise in PRO—a gain that outweighs the slight imprecision that results from using derivatives rather than difference equations.

Downward sloping. We assume that $d(p)$ is downward sloping whenever $d(p) > 0$—that is, that raising prices for a good during a period will decrease demand for that good during that period unless demand is already 0, in which case demand will remain at 0. Conversely, lowering prices can only increase demand.

The “downward sloping” assumption calls for a bit more discussion. First of all, it should be noted that downward-sloping demand curves do not mean that high prices will always be associated with low demand. A hotel revenue management will experience higher average rates when occupancy is high and lower average rates when occupancy is low. What the downward-sloping property does indicate is that, in any time period, demand would have been lower if prices had been higher, and vice versa. This corresponds both to economic theory (in which consumers maximize their utilities subject to a budget constraint) and to real-life experience.

Nonetheless, there are at least three cases in which the downward-sloping property may not hold.

1. *Giffen goods.* Economic theory allows for the possibility of so-called *Giffen goods*, whose demand rises as their price rises because of substitution effects. An example might be a student on a strict budget of $8.00 per week for dinner. When hamburger costs $1.00 per serving and steak costs $2.00, he eats hamburger six times a week and steak once. If the price of hamburger rises to $1.10, to stay within his budget, he stops buying steak and buys hamburger seven times a week. In this case, a rise in the price of hamburger causes his consumption of hamburger to increase. While this behavior might conceivably occur at an individual level, a Giffen good requires that enough buyers act this way that they overwhelm other buyers who would buy less hamburger as the price rises. Giffen goods are almost never encountered in reality—in fact, many economists doubt whether they have ever existed.
2. *Price as an indicator of quality.* In some markets, price is used by some consumers as an indicator of quality: Higher prices signal higher quality. In this case, lowering the price for a product may lead consumers to believe that it is of lower quality, and demand could drop as a result. Typically, markets where this is an issue have a large number of alternatives and some "lazy" buyers who do not have the time or resources to research the relative quality of all the alternatives so that they use price as a proxy. Wine is a classic example: Faced with a daunting array of labels and varietals, many purchasers are likely to use a rule such as: a $10 bottle for dinner with the family, a $15 bottle if the couple next door is dropping by, and a $25 bottle if our wine-snob friends are joining us for

dinner.6 The "price-as-an-indicator-of-quality effect" can be particularly important when a new product enters the market. A medical-product company developed a way of producing a home testing device at a cost 75% below the cost of the prevailing technology. They introduced the new product at a list price 60% lower than the list price of the leading competitors, expecting to dominate the market. When sales were slow, they repackaged the product and sold it at a price only 20% lower than the leading competitor. This time sales took off. Their belief is that the initial rock-bottom price induced customers to believe their product was inferior and unreliable. The higher price was high enough not to raise quality concerns but low enough to drive high sales.

3. *Conspicuous consumption*. Thorstein Veblen coined the term *conspicuous consumption* for the situation in which a consumer makes a purchase decision in order to advertise his ability to spend large amounts. It probably does not come as a shock that the reason some rock stars drink $300 bottles of Cristal champagne and drive Bentleys is not their finely honed appreciation of fine French champagne and British automotive engineering. Conspicuous consumption postulates a segment of customers who buy a product simply because it has a high price—and others know it. Dropping the price in this case may cause the product to lose its cachet and decrease demand.

While duly noting these three exceptions to downward-sloping demand curves, we will proceed to ignore them for the remainder of the book. In defense of this decision, we observe that for almost all items, almost all of the time, raising the price will lower demand and lowering the price will increase demand.

3.1.2 Measures of Price Sensitivity

It is often useful to have a simple characterization of the price sensitivity implied by a price-response function at a particular price. The two most common measures of price sensitivity are the *slope* and the *elasticity* of the price-response function.

Slope. The *slope* of the price-response function measures how demand changes in response to a price change. It is equal to the change in demand divided by the difference in prices, or

$$\delta(p_2, p_1) = [d(p_2) - d(p_1)]/(p_2 - p_1) \tag{3.1}$$

By the downward-sloping property, $p_1 > p_2$ implies that $d(p_1) \leq d(p_2)$. This means that $\delta(p_1,p_2)$ will always be less than or equal to zero.

The definition in Equation 3.1 requires two prices to be specified, because the slope of a price-response function will be constant across all prices only if it is linear. However, it is common to specify the slope at a single price, say, p_1, in which case it can be computed as the limit of Equation 3.1 as p_2 approaches p_1 That is,

$$\begin{aligned}\delta(p_1) &= \lim_{h \to 0}[d(p_1 + h) - d(p_1)]/h \\ &= d'(p_1)\end{aligned}$$

where $d'(p_1)$ denotes the derivative of the price-response function at p_1. By the differentiability property, we know that this derivative exists. The downward-sloping property means that the slope will be less than or equal to zero for all prices.

The slope can be used as a *local* estimator of the change in demand that would result from a small change in price. For small changes in price, we can write

$$d(p_2) - d(p_1) \approx \delta(p_1)(p_2 - p_1) \tag{3.2}$$

That is, a large (highly negative) slope means that demand is more responsive to price than a smaller (less negative) slope.

Example 3.1

The slope of the price-response function facing a semiconductor manufacturer at the current price of $0.13 per chip is 1,000 chips/week per cent. From Equation 3.2, he would estimate that a 2-cent increase in price would result in a reduction in demand of about 2,000 chips per week and a 3-cent decrease in price would result in approximately 3,000 chips/week in additional demand.

It is important to recognize that the quality of the approximation in Equation 3.2 declines for larger changes in prices and that the slope cannot be used as an accurate predictor of demand at prices far from the current price. It is also important to realize that the slope of the price-response function depends on the units of measurement being used for both price and demand.

Example 3.2

The price of a bulk chemical can be quoted in either cents per pound or dollars per ton. Assume that the demand for the chemical is 50,000 pounds at 10 cents per pound but drops to 40,000 pounds at 11 cents per pound. The slope of the price response function at these two points is

$$\begin{aligned}\delta(10, 11) &= (50{,}000 - 40{,}000)/(10 - 11)\\ &= -10{,}000 \text{ pounds/cent}\end{aligned}$$

The same slope in tons per dollar would be (25 20)/(0.1– 0.11) 500 tons/dollar.

Price elasticity. Perhaps the most common measure of the sensitivity of demand to price is *price elasticity*, defined as the ratio of the percentage change in demand to the percentage change in price.7 Formally, we can write

$$\epsilon(p_1, p_2) = -\frac{100\{[d(p_2) - d(p_1)]/d(p_1)\}}{100\{(p_2 - p_1)/p_1\}} \quad (3.3)$$

where $\varepsilon(p_1, p_2)$ is the elasticity of a price change from p_1 to p_2. The numerator in Equation 3.3 is the percentage change in demand, and the denominator is the percentage change in price. Reducing terms gives

$$\epsilon(p_1, p_2) = -\frac{[d(p_2) - d(p_1)]p_1}{[p_2 - p_1]d(p_1)} \quad (3.4)$$

The downward-sloping property guarantees that demand always changes in the opposite direction from price. Thus, the minus sign on the right-hand side of Equation 3.4 guarantees that $\varepsilon(p_1, p_2) \geq 0$. An elasticity of 1.2 means that a 10% *increase* in price would result in a 12% *decrease* in demand and an elasticity of 0.8 means that a 10% decrease in price would result in an 8% increase in demand.

$\varepsilon(p_1, p_2)$ as defined in Equation 3.4 is sometimes called the *arc elasticity*. That it requires two prices to calculate reflects that the percentage change in demand resulting from changing prices will depend on both the old price and the new price. In fact, the percentage decrease in from a 1%*increase* in price will generally not even be the same as the percentage increase in demand that we would experience from a 1% *decrease in price*. For this reason, both prices need to be specified in order to fully characterize elasticity. However, as we did with slope, we can derive a *point elasticity* at p by taking the limit of Equation 3.4 as p_2 approaches p_1:

$$\epsilon(p_1) = -d'(p_1)p_1/d(p_1) \quad (3.5)$$

In words, the point elasticity is equal to −1 times the slope of the demand curve times the price, divided by demand. Since $d'(p) \leq 0$, the point elasticity $\varepsilon(p)$ calculated by Equation 3.5 will be greater than or equal to zero. The point elasticity is useful as a local estimate of the change in demand resulting from a *small* change in price.

Example 3.3

A semiconductor manufacturer is selling 10,000 chips per month at $0.13 per chip. He believes that the price elasticity for his chips is 1.5. Thus, a 15% increase in price from $0.13 to $0.15 per chip would lead to a decrease in demand of about 1.5 15% 22.5%, or from 10,000 to about 7,750 chips per month.

One of the appealing properties of elasticity is that, unlike slope, its value is independent of the units being used. Thus, the elasticity of electricity is the same whether the quantity electricity is measured in kilowatts or megawatts and whether the units are dollars or euros.

Example 3.4

Consider the bulk chemical whose price-response slope was estimated in Example 3.2. It showed a 20% decrease in demand (from 50,000 pounds to 40,000 pounds) from a 10% increase in price (from 10 cents to 11 cents). The corresponding elasticity is 0.2/0.1 = 2—an elastic response. What if the units were euros and tons? It would still be a 20% decrease in demand (from 25 tons to 20 tons) from a 10% increase in price.

Like slope, point elasticity is a *local* property of the price-response function. That is, elasticity can be specified between two different prices by Equation 3.4 and a point elasticity can be defined by Equation 3.5. However, the term *price elasticity* is often used more broadly and somewhat loosely. Thus, statements such as "gasoline has a price elasticity of 1.22" are imprecise unless they specify both the time period of application and the reference price. In practice, the term *price elasticity* is often used simply as a synonym for *price sensitivity*. Items with "high price elasticity" have demand that is very sensitive to price while "low price elasticity" items have much lower sensitivity. Often, a good with a price elasticity greater than 1 is described as elastic, while one with an elasticity less than 1 is described as inelastic.

Elasticity depends on the time period under consideration, and, as with other aspects of price response, we must be clear to specify the time frame we are talking about. For most products, *short-run elasticity* is lower than *long-run elasticity*. The reason is that buyers have more flexibility to adjust to higher prices in the long run. For example, the short-run elasticity for gasoline has been estimated to be 0.2, while the long-run elasticity has been estimated at 0.7. In the short run, the only options consumers have in response to high gas prices are to take fewer trips and to use public transportation. But if gasoline prices stay high, consumers will start buying higher mile-per-gallon cars, depressing overall demand for gasoline even further. A retailer raising the price of milk by 20 cents may not see much change in milk sales for the first week or so and conclude that the price elasticity of milk is low. But he will likely see a much greater deterioration in demand over time. The reason is that customers who come to shop for milk after the price rise will still buy milk, since it is too much trouble to go to another store. But some customers will note the higher price and switch stores the next time they shop.

On the other hand, the long-run price elasticity of many durable goods—such as automobiles and washing machines—is lower than the short-run elasticity. The reason is that customers initially respond to a price rise by postponing the purchase of a new item. However, they will still purchase at some time in the future, so the long-run effect of the price change is less than the short-run effect.

Finally, it is important to specify the level at which we are calculating elasticity. Market elasticity measures total market response if all suppliers of a product increase their prices—perhaps in response to a common cost change. Market elasticity is generally much lower than the price-response elasticity faced by an individual supplier within the market. The reason is simple: If all suppliers raise their price, the only alternative faced by customers is to purchase a substitute product or to go without. On the other hand, if a single supplier raises its price, its customers have the option of defecting to the competition.

Table 3.1 shows some elasticities that have been estimated for various goods and services. Note that a staple such as salt is very inelastic—customers do not change the amount of salt they purchase very much in response to *market* price changes. On the other hand, we would expect that price elasticity of the market-response function faced by any individual seller of salt to be quite large—since salt is a fungible commodity in a highly competitive market. This effect can be seen in Table 3.1 in the difference between the short-run elasticity for automobile purchases (1.2) and the much larger elasticity (4.0) faced by Chevrolet models. The table also illustrates the fact that long-run elasticity is greater than the short-run elasticity for airline travel (where customers respond to a price rise by changing travel plans in the future and traveling less by plane), but the reverse is true for automobiles (where consumers respond to price rises by postponing purchases).

TABLE 3.1

Estimated price elasticities for various goods and services

Good	Short-run elasticity	Long-run elasticity
Salt	0.1	—
Airline travel	0.1	2.4
Tires	0.9	1.2
Restaurant meals	2.3	—
Automobiles	1.2	0.2
Chevrolets	4.0	

3.1.3 Price Response and Willingness to Pay

So far, we have treated the price-response function as simply given. In reality, demand is the result of thousands, perhaps millions, of individual buying decisions on the part of potential customers. Each potential customer observes our price and decides whether or not to buy our product. Those who do not buy our product may purchase from the competition, or they may decide to do without. The price-response function specifies how many more of those potential customers would buy if we lowered our price and how many current buyers would not buy if we raised our price. Thus the price-response function is based on assumptions about customer behavior. We usually cannot directly track the thousands or even millions of individual decisions that ultimately manifest themselves in demand for our product.[8] However, it is worthwhile to understand the assumptions about customer behavior that underlie the price-response functions so that we can judge if the price-response function is based on assumptions appropriate for the application. The most important of such models of customer behavior is based on *willingness to pay*.

The willingness-to-pay approach assumes that each potential customer has a *maximum willingness to pay* (sometimes called a *reservation price*) for a product or service. A customer will purchase if and only if the price is less than her maximum willingness to pay. (We will use *willingness to pay*, sometimes abbreviated *w.t.p.*, to mean "maximum willingness to pay.") For example, a customer with a willingness to pay of $253 for an airline ticket from New

York to Miami will purchase the ticket if the price is less than or equal to $253 but not if it is $253.01 or more. In this case, $d(253)$ equals the number of customers whose maximum willingness to pay is at least $253. A customer with a maximum willingness to pay of $0 (or less) will not buy at any price.

Define the function $w(x)$ as the w.t.p. distribution across the population. Then, for any values $0 \leq p_1\, p_2$:

$$\int_{p_1}^{p_2} w(x)\, dx = \text{fraction of the population that has w.t.p. between } p_1 \text{ and } p_2$$

We note that $0 \leq w(x) \leq 1$ for all nonnegative values of x. Let $D = d(0)$, the maximum demand achievable. Then we can derive $d(p)$ from the w.t.p. distribution from

$$d(p) = D\int_{p}^{\infty} w(x)\, dx \tag{3.6}$$

We can take the derivative of the corresponding price-response function to obtain

$$d'(p) = -Dw(p)$$

which is nonpositive, as required by the downward-sloping demand curve property. Conversely, we can derive the willingness-to-pay distribution from the price-response function[9] using

$$w(x) = -d'(x)/d(0)$$

Example 3.5

The total potential market for a spiral-bound notebook is $D = 20{,}000$, and willingness to pay is distributed uniformly between $0 and $10.00 as shown in Figure 3.3. This means that

$$w(x) = \begin{cases} 1/10 & \text{if } 0 \leq x \leq \$10 \\ 0 & \text{otherwise} \end{cases}$$

We can apply Equation 3.6 to derive the corresponding price-response curve:

$$\begin{aligned} d(p) &= 20{,}000 \int_p^{10} (1/10)\, dx \\ &= 20{,}000(1 - p/10) \\ &= 20{,}000 - 2{,}000p \end{aligned}$$

The price-response curve $d(p) = 20{,}000 - 2{,}000p$ is a straight line with $d(0) = 20{,}000$ and a satiating price of \$10.00.

Example 3.5 illustrates a general principle:

A uniform willingness-to-pay distribution corresponds to a linear price-response function, and vice versa.

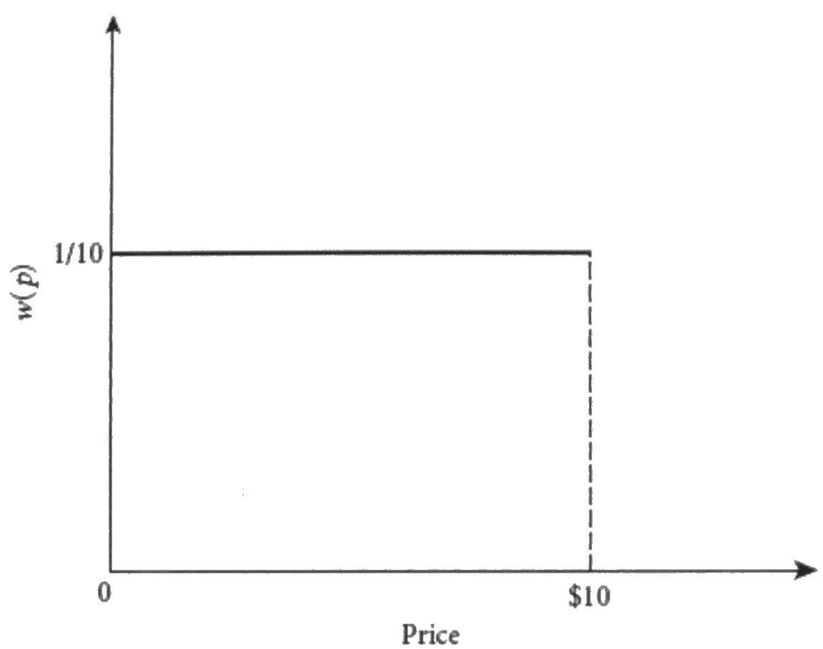

Figure 3.3 Uniform willingness-to-pay distribution.

One of the advantages of Equation 3.6 is that it partitions the price-response function into a total-demand component D and a willingness-to-pay component $w(x)$. This is often a convenient way to model a market. For example, we might anticipate that total demand varies seasonally for some product while the willingness-to-pay distribution remains constant over time. Then, given a forecast of total demand $D(t)$ for a future period t, our expected price-response function in each future period would be

$$d(p(t)) = D(t)\int_{p}^{P} w(x)\, dx \tag{3.7}$$

This approach allows us to decompose the problem of forecasting total demand from the problem of estimating price response. It also allows us to model influences on willingness to pay and total demand independently and then to combine them. For example, we might anticipate that a targeted advertising campaign will not increase the total population of potential customers, $D(t)$, but that it will shift the willingness-to-pay distribution. On the other hand, if we open a new retail outlet, we might anticipate that the total demand potential for the new outlet will be determined by the size of the population served, while the willingness to pay will have the same distribution as existing stores serving populations with similar demographics.

Of course, a customer's willingness to pay changes with changing circumstances and tastes. The maximum willingness to pay for a cold soft drink increases as the weather gets warmer—a fact that the Coca-Cola company considered exploiting with vending machines that changed prices with temperature (see Chapter 12 for a discussion of the "temperature-sensitive vending machine" idea). Willingness to pay to see a movie is higher for most people on Friday night than on Tuesday afternoon. A sudden windfall or a big raise may increase an individual's maximum willingness to pay for a new Mercedes Benz. To the extent that such changes are random and uncorrelated, they will not effect the overall w.t.p. distribution, since increasing willingness to pay on one person's part will tend to be balanced by another's decreasing willingness to pay. On the other hand, systematic changes across a population of customers will change the overall distribution and cause the price-response function to shift. Such systematic changes may be due to seasonal effects, changing fashion or fads, or an overall rise in purchasing power for a segment of the population. These systematic changes need to be understood and incorporated into estimating price response and future price response.

A disadvantage of the willingness-to-pay formulation is that it assumes that customers are considering only a single purchase. This is a reasonable assumption for relatively expensive and durable items. However, for many inexpensive or nondurable items, a reduction in price might cause some customers to buy multiple units. A significant price reduction on a washing machine will induce additional customers to buy a new washing machine, but it is unlikely to induce many customers to purchase two. However, a deep discount on socks may well induce customers to buy several pairs. This additional induced demand is not easily incorporated in a willingness-to-pay framework—willingness-to-pay models are most applicable to "big ticket" consumer items and industrial goods.

3.1.4 Common Price-Response Functions

Linear price-response function. We have seen that a uniform distribution of willingness to pay generates a linear price-response function. The general formula for the linear price-response function is

$$d(p) = D - mp \tag{3.8}$$

where $D > 0$ and $m > 0$. $D = d(0)$ is the *demand at zero price*. The general linear price-response function is shown in Figure 3.4. The *satiating price*—that is, the price at which demand drops to zero—is given by $P = D/m$. The slope of the linear price-response function is m for 0 p P and 0 for $p \geq P$. The elasticity of the linear price-response function is $mp/(D - mp)$, which ranges from 0 at $p = 0$ and approaches infinity as p approaches P, dropping again to 0 for $p >$ P.

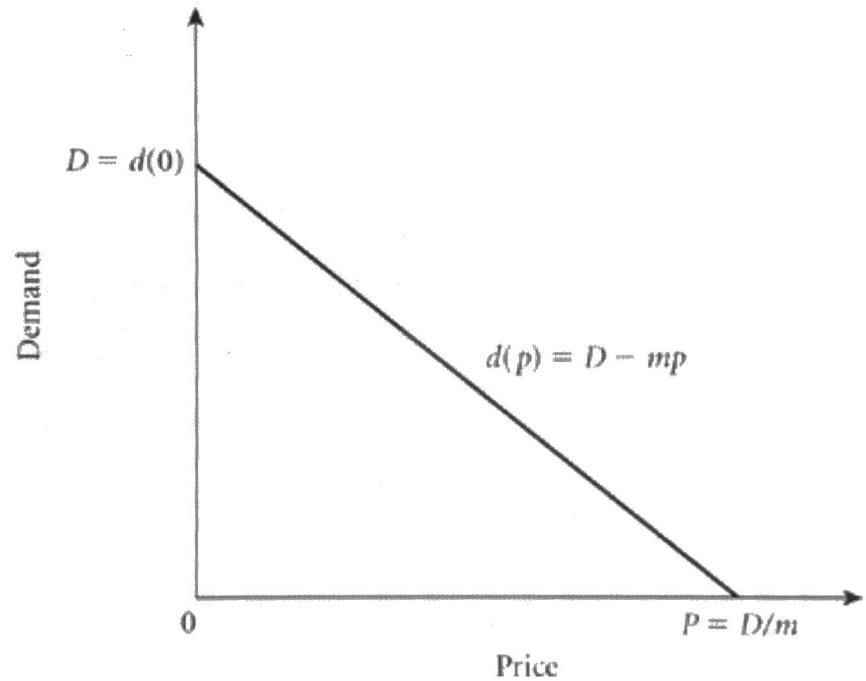

Figure 3.4 Linear price-response function.

We will use the linear price-response function in many examples because it is a convenient and easily tractable model of market response. However, it is not a realistic *global* representation of price response. The linear price-response function assumes that the change in demand from a 10-cent increase in price will be the same, no matter what the base price might be. This is unrealistic, especially when a competitor may be offering a close substitute. In this case, we would usually expect the effect of a price change to be greatest when the base price is close to the competitor's price.

Constant-elasticity price-response function. As the name implies, the constant-elasticity price-response function has a point elasticity that is the same at all prices. That is,

$$d'(p)p/d(p) = -\epsilon \qquad \text{for all } p > 0 \tag{3.9}$$

where $\varepsilon > 0$ is the elasticity. The price-response function corresponding to Equation 3.9 is

$$d(p) = Cp^{-\epsilon} \tag{3.10}$$

where $C > 0$ is a parameter chosen such that $d(1) = C$. For example, if we are measuring price in dollars and $d(\$1.00) = 10{,}000$, then $C = 10{,}000$. The slope of the constant-elasticity price-response function is

$$d'(p) = -C\varepsilon p^{-(\varepsilon+1)}$$

which is less than zero, confirming that constant-elasticity price-response functions are downward sloping. Some examples of constant-elasticity price-response functions are shown in Figure 3.5. Note that constant-elasticity price-response functions are neither finite nor satiating. Demand does not drop to zero at any price, no matter how high, and demand continues to approach infinity as price approaches zero. For these reasons, constant elasticity is usually not a good *global* assumption for price response.

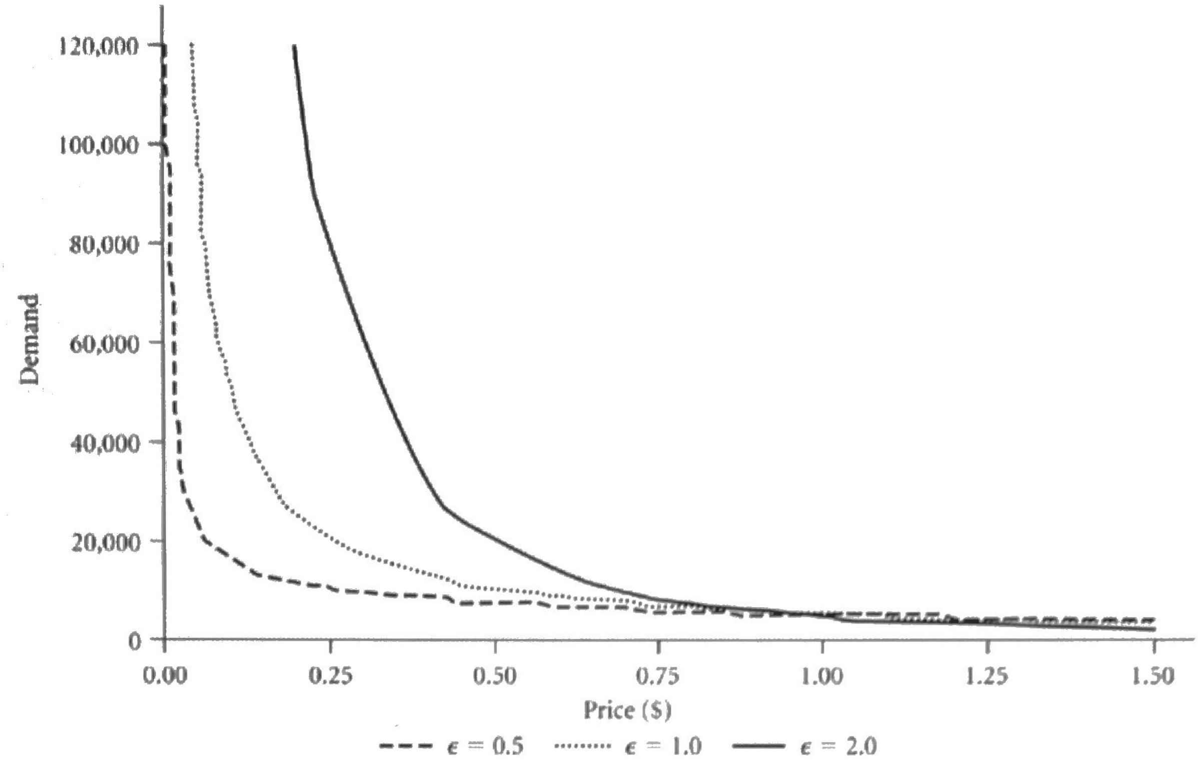

Figure 3.5 Constant-elasticity price-response functions.

It is interesting to analyze the impact of price on *revenue* for a seller facing constant-elasticity price response. Let $R(p)$ denote total revenue at price p. Then $R(p) = p\,d(p)$ and, for the constant-elasticity price-response function given by Equation 3.10,

$$R(p) = pCp^{-\varepsilon}$$

$$= Cp^{(1-\varepsilon)}$$

Taking the slope of this function gives

$$R'(p) = (1 - \varepsilon)Cp^{-\varepsilon}$$

$$= (1 - \varepsilon)\, d(p)$$

Since $d(p) > 0$, the direction of the slope is determined by the $(1 - \varepsilon)$ term. If ε 1 (that is, if demand is inelastic), then $R'(p) > 0$. This means that the seller facing constant *inelastic* price response can always increase his revenue by increasing price. In fact, he can also increase total operating margin by increasing price, so we would never expect to find a company maximizing operating margin setting its price such that the local elasticity is less than 1. If $\varepsilon > 1$, a seller can increase revenue by decreasing price. In fact, the seller can maximize revenue by setting the price as close to zero—without actually becoming zero—as possible! If $\varepsilon = 1$, then $R(p) = C$ for any value of p—revenue will not change at all as price changes. Revenue as a function of price for three different constant-elasticity price-response cases is illustrated in Figure 3.6.

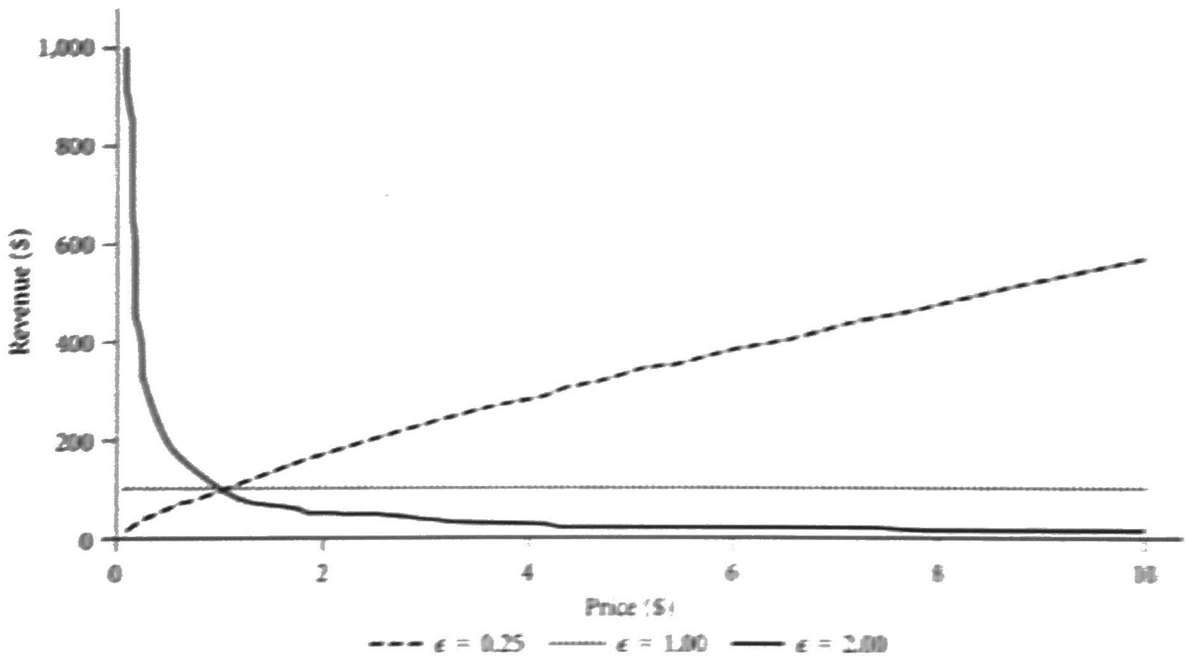

Figure 3.6 Revenue as a function of price for the constant-elasticity price-response function.

These properties make the constant-elasticity function problematic as a global price-response function. It is certainly the rare market in which a seller can maximize his revenue by either increasing his price toward infinity or dropping it as close to zero as possible. For these reasons, constant elasticity is a better representation of local price response than for use as a global price-response function. In real-world situations, we would expect price elasticity to change as price changes.

The constant-elasticity price-response function in Equation 3.10 corresponds to the willingness-to-pay distribution given by

$$w(x) = \varepsilon x^{-(\varepsilon + 1)}$$

This distribution is given in Figure 3.7 for different values of ε. The figure shows that the constant-elasticity price-response function corresponds to a distribution of willingness to pay that is highly concentrated near zero.

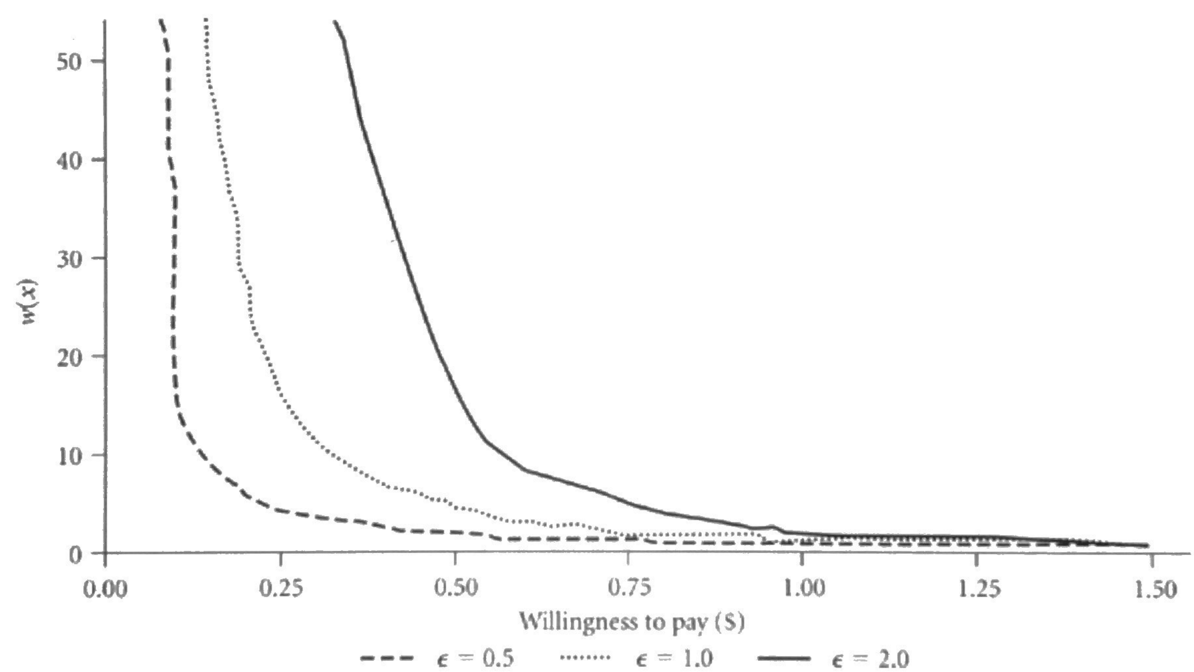

Figure 3.7 Constant-elasticity willingness-to-pay distributions.

The logit price-response functions. Both the linear and the constant-elasticity price-response functions are useful local models of price response. However, they have rather severe limitations as global models of customer behavior. This can be seen by considering their corresponding willingness-to-pay distributions. The linear price-response function assumes that w.t.p. is uniformly distributed between 0 and some maximum value. The constant-elasticity price-response function assumes that the distribution of willingness to pay drops steadily as price increases, approaching, but never reaching, zero. Neither of these functions would seem to represent customer behavior very realistically.

How would we expect customers to react to changes in price? Consider the case where we are selling a compact car. Assume that competing models are generally similar to ours and are selling at a market price of about $13,000. If we price well above the market price, say, at $20,000, we won't sell many cars. But the few customers who are buying from us at $20,000 must be very loyal (or very ignorant)—we would not expect to lose very many more of them if we increased our price from $20,000 to $21,000. In other words, the elasticity of the price-response function is low at this price. On the other hand, if we priced well below the market price—say, at $9,000—we would expect to sell lots of cars. At this price, almost anyone who wants to buy a compact car and is not incredibly loyal to another brand (or incredibly averse to ours) would be buying from us. Lowering our price another $1,000 is unlikely to attract many new customers. Again, price elasticity is low.

What happens when we are close to the market price? In this case, we would expect small changes in our price to lead to substantial shifts in demand. There are a lot of customers who are more or less indifferent between our offering and that of the competition. Many of these customers can be persuaded to purchase our car if we ask only $250 less than the market price, and even more will shift if we ask $500 less than the market price. Similarly, if we charge $250 above the market price we are likely to lose a lot of demand to the competition. In other words, price elasticity is highest when we are at (or near) the market price.

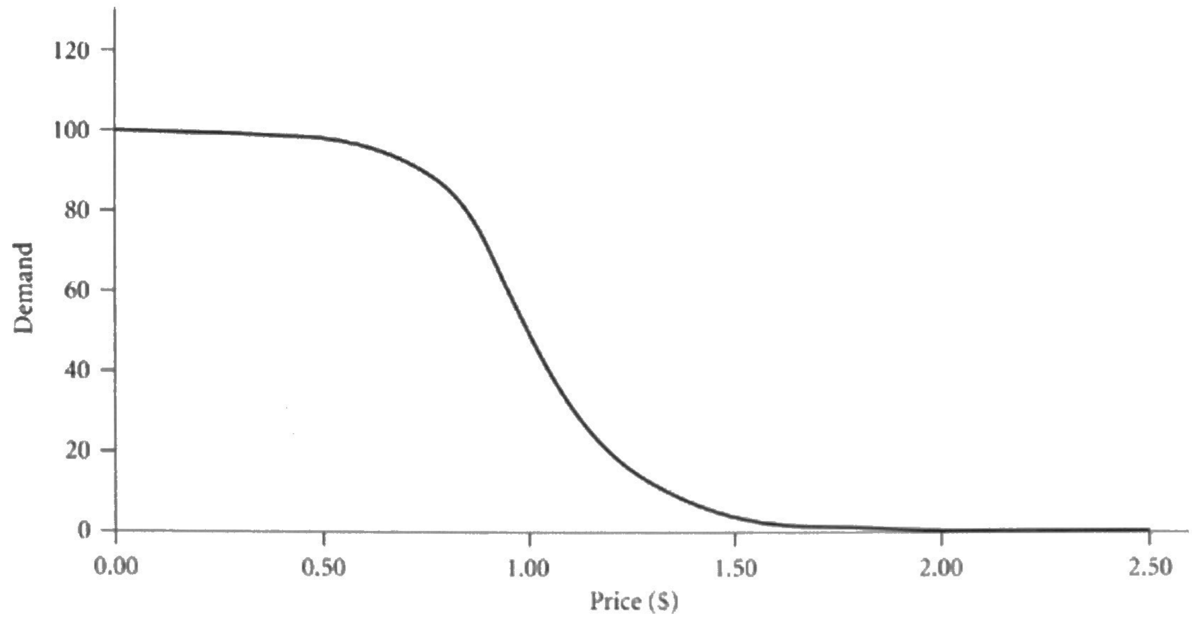

Figure 3.8 (Reverse) S-shaped, or sigmoid, price-response curve.

This kind of consumer behavior generates a response curve of the general form shown in Figure 3.8. When we price very low, we receive lots of demand, but demand changes slowly as we change price. In the area of the "market price," demand is very sensitive to our price—small changes in price can lead to substantial changes in demand. At high prices, demand is low and changes slowly as we raise prices further. The price-response curve has a sort of reverse S shape. Empirical research has shown that this general form of price-response curve fits observed price response in a wide range of markets.

The most popular price-response function of this form is the *logit* price-response function:

$$d(p) = \frac{Ce^{-(a+bp)}}{1+e^{-(a+bp)}} \tag{3.11}$$

Here, a, b, and C are parameters with $C > 0$ and $b > 0$. a can be either greater than or less than 0, but in most applications we will have $a > 0$. Broadly speaking, C indicates the size of the overall market and b specifies price sensitivity. Larger values of b correspond to greater price sensitivity. The price-sensitivity curve is steepest at the point *p*^ = -(*a*/*b*), as indicated in Figure 3.8. This point can be considered to be approximately the "market price."

The logit price-response curve is shown for different values of *b* in Figure 3.9. Here we have fixed *p*^ = $13,000, so *a* = -$13,000 × *b* for each curve. Demand is very sensitive to price when price is close to *p*^. Higher values of *b* represent more price-sensitive markets. As *b* grows larger, the market approaches perfect competition. In other words, the price-response curve increasingly approaches the perfectly competitive price-response function in Figure 3.1.

Some of the characteristics of the logit price-response function are shown in Table 3.2. Logit willingness to pay follows a bell-shaped curve known as the *logistic distribution*. The logistic distribution is similar to the normal distribution, except it has somewhat "fatter tails"—that is, it approaches zero more slowly at very high and very low values. An example of the logistic w.t.p. distribution is shown in Figure 3.10. The highest point (mode) of the logistic willingness-to-pay distribution occurs at *p*^ = -(*a*/*b*), which is also the point at which the slope of the price response is steepest. This is a far more realistic w.t.p. distribution than those associated with either the constant-elasticity or linear price-response functions. For that reason, the logit is usually preferred to linear or constant-elasticity price-response functions when the effects of large price changes are being considered.[10]

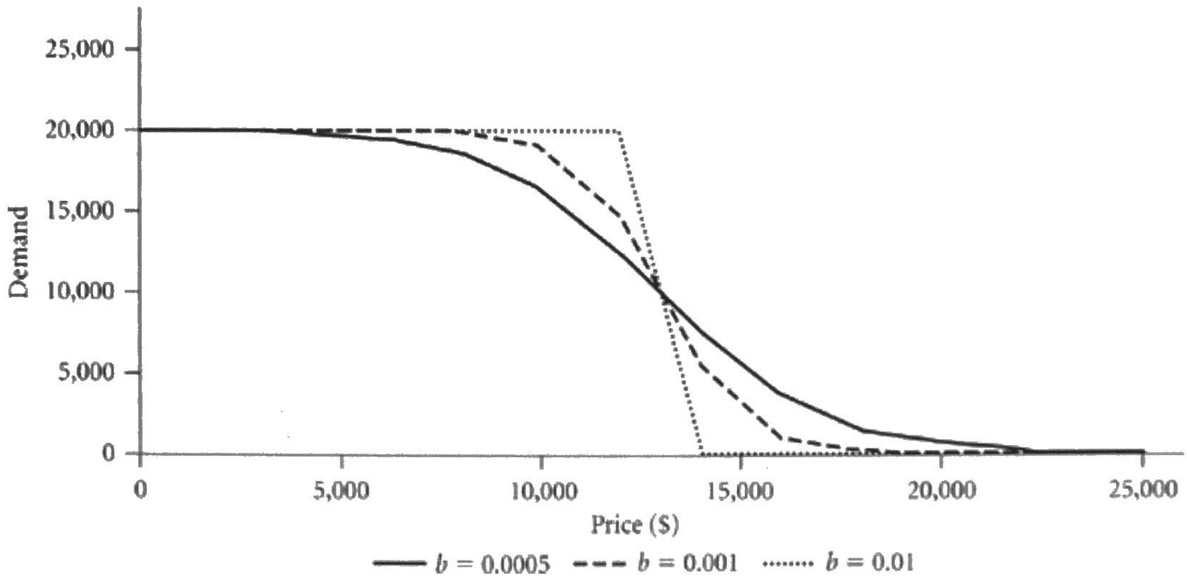

Figure 3.9 Logit price-response functions with *C* 20,000 and *p* $13,000.

TABLE 3.2
Properties of the logit price-response function

Demand at 0	$d(0) = \dfrac{Ce^{-a}}{1 + e^{-a}}$
Slope	$d'(p) = \dfrac{-Cbe^{-(a+bp)}}{(1 + e^{-(a+bp)})^2}$
Elasticity	$\epsilon(p) = \dfrac{bp}{1 + e^{-(a+bp)}}$
Willingness-to-pay distribution	$w(x) = \dfrac{Ke^{-(a+bx)}}{(1 + e^{-(a+bx)})^2}$

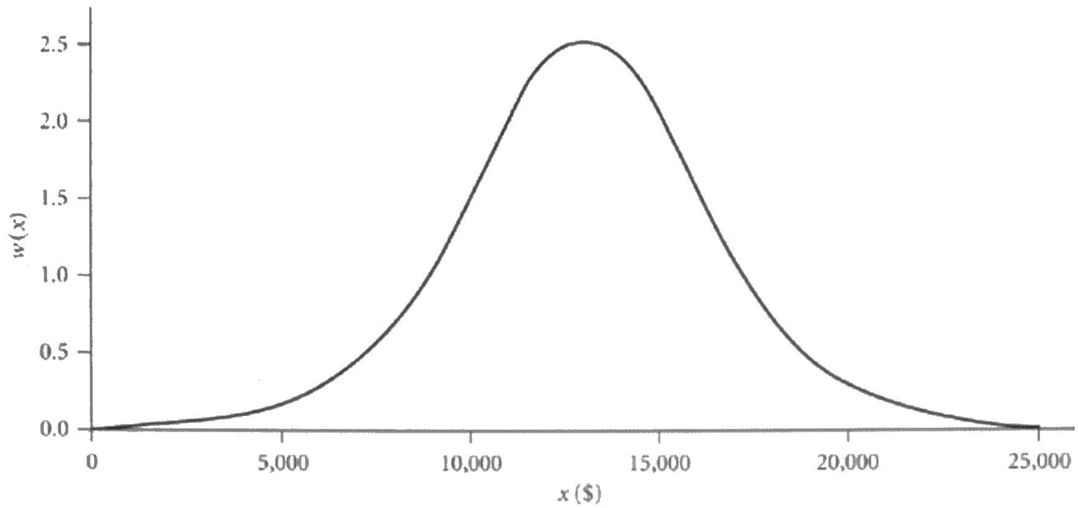

Figure 3.10 Willingness-to-pay distribution corresponding to the logit price-response function in Figure 3.9 with b = 0.0005. The y-axis has been scaled by a factor of 20,000.

3.2 PRICE RESPONSE WITH COMPETITION

So far we have not given much attention to the issue of competition. We have mentioned a "market price" as the steepest point in the logit price-response function, but we have not made an explicit link with competition. Yet competition is an important—some may say the most important—fact of life in any market of interest. Most managers identify competition as the number-one factor influencing their pricing. Many complain about being at the mercy of their competitors who have "stupid" or "irrational" pricing. How should we factor competition into our calculation of price response?

There are three different levels at which competition might be included in PRO. In increasing levels of sophistication and difficulty they are: incorporating competition in the price-response function, explicitly modeling consumer choice, and trying to anticipate competitive reaction.

3.2.1 Incorporating Competition in the Price-Response Function

part on what portion of the customer's freight is *backhaul* business, which can utilize excess capacity on existing trucks, and what portion will be *headhaul* freight, which will require running new trucks.

In each case, the calculation of incremental cost requires understanding the nature of the customer commitment and then estimating the additional costs that would be generated by making the commitment—or, equivalently, the costs that would be avoided by not making the commitment. The methodology behind calculating incremental costs is closely related to activity-based costing (ABC). Activity-based costing is a management accounting approach to allocating costs to their underlying causes in order to give a clearer view of the real sources of cost within an organization. A good introduction to activity-based costing is *Cost & Effect* by Kaplan and Cooper (1997).

3.4 THE BASIC PRICE OPTIMIZATION PROBLEM

The difference between the price at which a product is sold and its incremental cost is called its *unit margin* or just *margin*. The sum of the margins of all products sold during a time period is called the *total contribution*. In most cases, the seller's goal is to maximize total contribution. When the supplier is selling a single product at a single price, her total contribution will be

$$m(p) = (p - c)\, d(p) \tag{3.15}$$

where $m(p)$ is total contribution and c is incremental cost. The basic price optimization problem is

$$\max_p (p - c)\, d(p) \tag{3.16}$$

The total contribution function $m(p)$ is hill shaped, with a single peak, as shown in Figure 3.11. The top of this peak is the maximum total contribution the supplier can realize in the current time period, and p^* is the price that will maximize the total contribution. Note that there is a fundamental lack of symmetry in the total-contribution curve: The supplier can lose money by pricing too low (below incremental cost), but she cannot lose money by pricing too high—the worst that can happen is that she drives demand to zero.

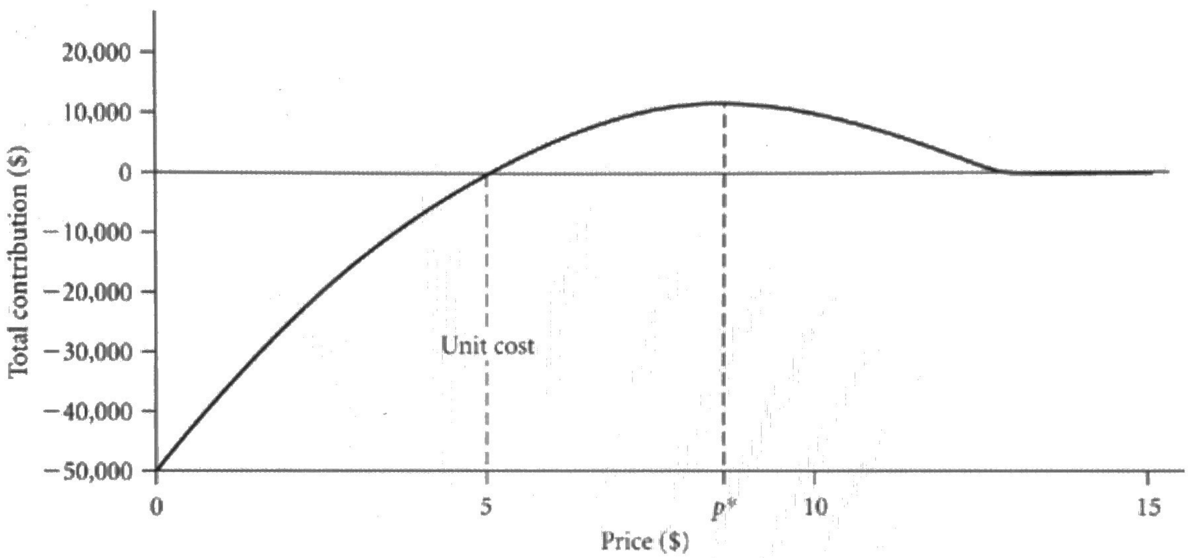

Figure 3.11 Total contribution as a function of price.

3.4.1 Optimality Conditions

The problem in Equation 3.16 is an unconstrained optimization problem, and standard optimization theory tells us that it can be solved by taking the derivative of $m(p)$ and setting it equal to zero.13 The derivative of $m(p)$ with respect to price is given by

$$m'(p) = d'(p)(p - c) + d(p) \tag{3.17}$$

To find the price that maximizes total contribution, we set $m'(p) = 0$, or

$$d'(p)(p - c) + d(p) = 0$$

Thus, a p^* satisfying

$$d(p^*) = -d'(p^*)(p^* - c) \tag{3.18}$$

will maximize total contribution.14

Condition 3.18 is not particularly insightful in itself, but it can be used to derive two standard conditions for optimal prices.

Marginal price equals marginal cost. We can rewrite Equation 3.18 as

$$p^*\, d'(p^*) + d(p^*) = c\, d'(p^*) \tag{3.19}$$

The term on the left-hand side of Equation 3.19 is *marginal revenue*—the derivative of total revenue with respect to price. This is the amount of additional revenue the seller could achieve from a small increase in price. Typically, marginal revenue is greater than zero at low prices but less than zero at higher prices. When price is low, increasing price leads to increased total revenue because the reduced demand is outweighed by increased margin. But at some price, the effect of raising price further is to decrease total revenue as demand begins to drop more quickly than margin increases.

The term on the right-hand side of Equation 3.19 is *marginal cost*: the amount of additional cost the seller would incur from a small increase in price. Note that marginal cost is always less than or equal to zero—an increase in price results in lower demand (by the downward-sloping property), which in turn leads to lower total costs. Equation 3.19 states that contribution is maximized when marginal revenue equals marginal cost.

Example 3.8

The marginal-revenue and marginal-cost curves are shown in Figure 3.12 for a seller with a price-response function given by $d(p) = 10{,}000 - 800p$ and incremental cost of \$5.00. The marginal-revenue curve corresponding to this price-response function is $R'(p) = 10{,}000 - 1{,}600p$, and the marginal cost curve is a horizontal line at $-\$4{,}000$. The contribution maximizing price occurs where the marginal-revenue curve intersects the marginal-cost curve in Figure 3.12, in this case at $p^* = \$8.75$.

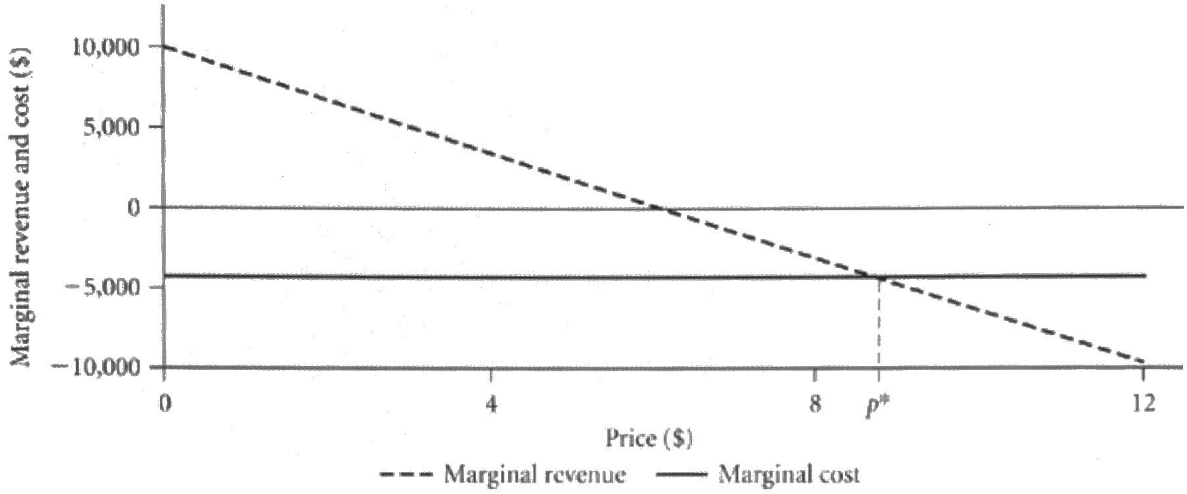

Figure 3.12 Marginal revenue and marginal cost.

We can state this important condition as follows.

> *Total contribution is maximized in the basic price optimization problem at the price at which marginal revenue equals marginal cost.*

Equation 3.19 provides further useful guidance on price changes that could improve total contribution. If marginal revenue is greater than marginal cost, then the supplier can increase his contribution by increasing price. If, on the other hand, marginal revenue is lower than marginal cost, he should decrease his price to increase contribution.

Optimal contribution margin and elasticity. We can also relate the optimal price to point elasticity. Rewrite Equation 3.17 as

$$\begin{aligned} m'(p) &= d(p)\left[\frac{d'(p)p}{d(p)} + 1\right] - d'(p)c \\ &= d(p)[1 - \epsilon(p)] - d'(p)c \qquad (3.20) \end{aligned}$$

where $\varepsilon(p)$ is point elasticity as defined in Equation 3.5. In Equation 3.20, the second term, $-d'(p)c$, will always be greater than or equal to zero since $d'(p) \leq 0$. If $d(p) > 0$, then $m'(p)$ will always be greater than zero if $\varepsilon(p)$ 1. In other words,

> *If the point elasticity at our current price is less than 1, we can increase total contribution by increasing price.*

Of course, since point elasticity changes as we change price, we cannot expect total contribution to continue increasing forever as we increase price. Typically, as price increases, elasticity will increase as well, until we reach a point where lost sales outweigh increased unit margins. We can express the corresponding condition in terms of point elasticity by combining Equation 3.20 with the condition that $m'(p^*) = 0$ for p^* to maximize total contribution. Then

$$\begin{aligned} d(p^*) &= -d'(p^*)(p^* - c) \\ -d'(p^*)/d(p^*) &= 1/(p^* - c) \\ -d'(p^*)p^*/d(p^*) &= p^*/(p^* - c) \\ \epsilon(p^*) &= p^*/(p^* - c) \end{aligned} \tag{3.21}$$

The quantity $(p - c)/p$ is the margin per unit expressed as a fraction of price. It is known as the *contribution margin ratio* or sometimes as the *gross margin ratio*. For a retailer purchasing an item at \$150 and selling it at \$200, the contribution margin ratio is (\$200 − \$150)/\$150 = 0.33. In words, Equation 3.21 says

> *At the optimal price, the price elasticity is equal to the reciprocal of the contribution margin ratio.*

Of course this is equivalent to

> *At the optimal price, the contribution margin ratio is equal to the reciprocal of*

elasticity.

A convenient formula relating the optimal price to elasticity and cost is

$$p^* = \frac{\epsilon(p^*)}{\epsilon(p^*) - 1} \times c$$

This relationship provides a particularly convenient way to calculate the optimal price in the face of a constant-elasticity price-response function.

Example 3.9

An electronics goods retailer faces a constant-elasticity price-response function with an elasticity of 2.5 for a popular model of television. It costs him \$180 apiece to purchase the televisions wholesale. At the optimum price, he should achieve a contribution margin ratio of 1/2.5 = 40% per unit. This means he should price the televisions at (2.5/1.5) × \$180 = \$300 in order to maximize total contribution.

Imputed price elasticity. Equation 3.21 implies that we can derive local price elasticity in a market from the contribution margin at the optimal price.

Example 3.10

A seller believes he is pricing optimally, and his contribution margin ratio is 20%. This can only be true if the price elasticity is 5.

Imputed price elasticity is a good “reality check” on the credibility of a company’s current pricing.

3.4.2 Applying Basic Price Optimization

We can illustrate basic price optimization with a simple example.

Example 3.11

A widget maker is looking to set the price of widgets for the current month. Assume that the widget maker's unit production cost c is a constant $5.00 per widget and that his demand for the current month is governed by the linear price-response function

$$d(p) = (10{,}000 - 800p)+$$

This means that the demand for widgets will be $10{,}000 - 800p$ for prices between zero and $12.50 and that the demand will be 0 for prices over $12.50. (We use the notation $(x)^+$ to denote the maximum of x and 0, where x may be either a single variable or a mathematical expression.) This linear price-response function is shown in Figure 3.13.

In this simple example, $d'(p) = 800$. Substituting into Equation 3.18, we can see that the optimal price p^* must solve

$$\$10{,}000\ 800 - 800p^* = 800(p^* - \$5.00)$$

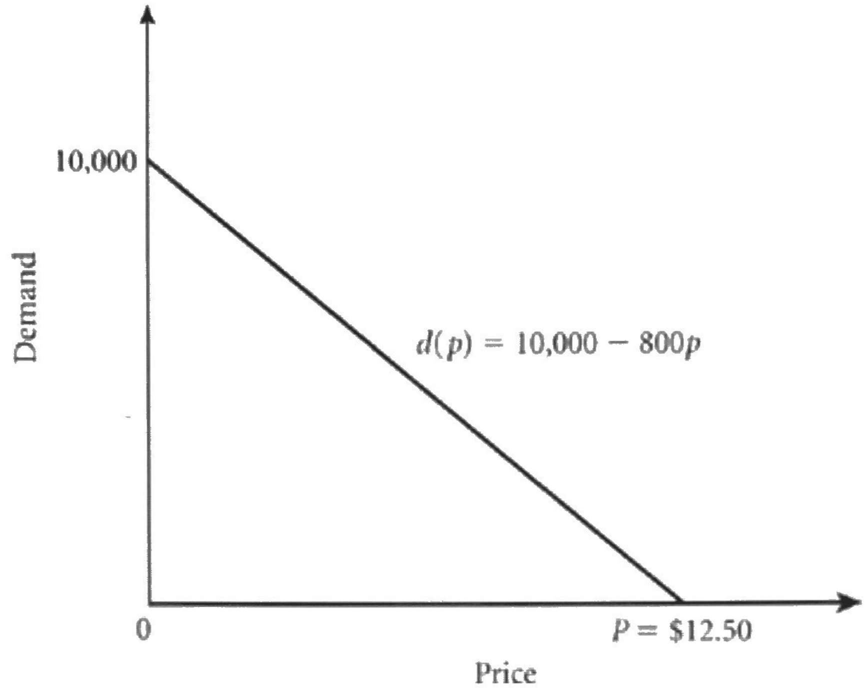

Figure 3.13 Price-response curve for widgets.

or, after a little algebra,

$$1{,}600p^* = 14{,}000 \quad \text{or} \quad p^* = \$8.75$$

At the optimal price of \$8.75, total widget sales will be 10,000 − 800(\$8.75) = 3,000 units, total revenue will be 3,000 × \$8.75 = \$26,250, and total contribution will be 3,000(\$8.75 − \$5.00), or \$11,250. Figure 3.11 shows how total contribution varies as a function of price. Note that total contribution is zero at $p = c =$ \$5.00, rises to a maximum at $p^* =$ \$8.75, and drops to zero again at $p =$ \$12.50.

3.4.3 Marginality Test

One way to check the optimality of p^* is to use the *marginality test*, which states that a particular price can only be optimal if raising the price by a penny or lowering the price by a penny results in reduced margin contribution. The principle of marginality should be pretty obvious—if we could increase contribution by changing the price, then the current price would not be optimal—but it is a useful check nonetheless. Table 3.3 shows the results of varying the price in Example 3.11 from $8.74 to $8.76. As expected, the price of $8.75 results in the highest margin of the three alternatives. However, the pattern of change among the prices is instructive: At $8.74 we are selling eight more units per month, but this is not enough to make up for the lost margin per unit. At $8.76 we are making a penny more per unit sold, but this is offset by the loss of sales of eight units. The optimal price, $8.75, exactly balances the gain in units sold from the potential loss of margin from raising prices further.

TABLE 3.3

Impact of price on margin contribution near the optimal price

Price	Demand	Unit margin	Margin contribution
$8.74	3008	$3.74	$11,249.92
$8.75	3000	$3.75	$11,250.00
$8.76	2992	$3.76	$11,249.92

3.4.4 Maximizing Revenue

In some cases, a company might wish to maximize total revenue (rather than total contribution). In this case, the objective function will be

$$\max_{p} R(p) = d(p)p \tag{3.22}$$

It is easy to see that this is equivalent to Equation 3.16 when $c = 0$. A company with incremental cost of zero can maximize net contribution by maximizing revenue. There are some service industries, such as movie theaters, video rentals, and sporting events, in which the incremental costs are close to zero. Some of these are discussed in Section 6.7.1. Also, incremental costs are zero (or very small) in many cases in which a company has already purchased a fixed amount of perishable, nonreplenishable inventory. This is the situation faced by many fashion-goods retailers, who purchase inventory for an entire season ahead of time. Once the inventory has been purchased, the incremental cost of a sale is zero—and the seller should set prices accordingly. Many of these situations count as *markdown opportunities* and are discussed in Chapter 10.

The revenue-maximizing price, p', can be found by differentiating $R(p)$ and setting the derivative equal to 0. Specifically,

$$R'(p') = d'(p')p' + d(p') = 0 \tag{3.23}$$

implying that p' solves

$$-\frac{d'(p')p'}{d(p')} = \epsilon(p') = 1 \tag{3.24}$$

In other words, the revenue-maximizing price occurs where the elasticity of the price-response function is equal to 1.[15] This implies that there is no unconstrained revenue-maximizing price associated with a constant-elasticity price-response function unless the price elasticity happens to be exactly 1—in which case, revenue is constant at all prices.

Equation 3.19 says that contribution is maximized when marginal revenue is equal to marginal cost (which is less than zero). Equation 3.23 says that contribution is maximized when marginal revenue is equal to zero. Typically, marginal revenue is a decreasing function of price (at least in the region of the optimal price). In this case, we can show how marginal revenue and marginal cost can be used to compute the revenue-maximizing and contribution-maximizing prices in Figure 3.14. Specifically, we can see that, as long as the marginal revenue curve is decreasing, *the revenue-maximizing price is lower than the contribution-maximizing price.*

Example 3.12

The CEO of the widget-making company decides that the firm's goal for the next month will be to maximize revenue from widget sales as part of the long-term strategy to increase market share. The revenue-maximizing price can be found by solving Equation 3.23. The resulting revenue-maximizing price is equal to $6.25, with corresponding sales of 5,000 units. The corresponding per-unit margin is $1.25, and the total contribution margin is 5,000 × $1.25 = $6,250. We can compare this to the maximum margin contribution of $11,250 and conclude that to maximize total revenue, the company needs to give up $11,250 - $6,250 = $5,000 of contribution margin.

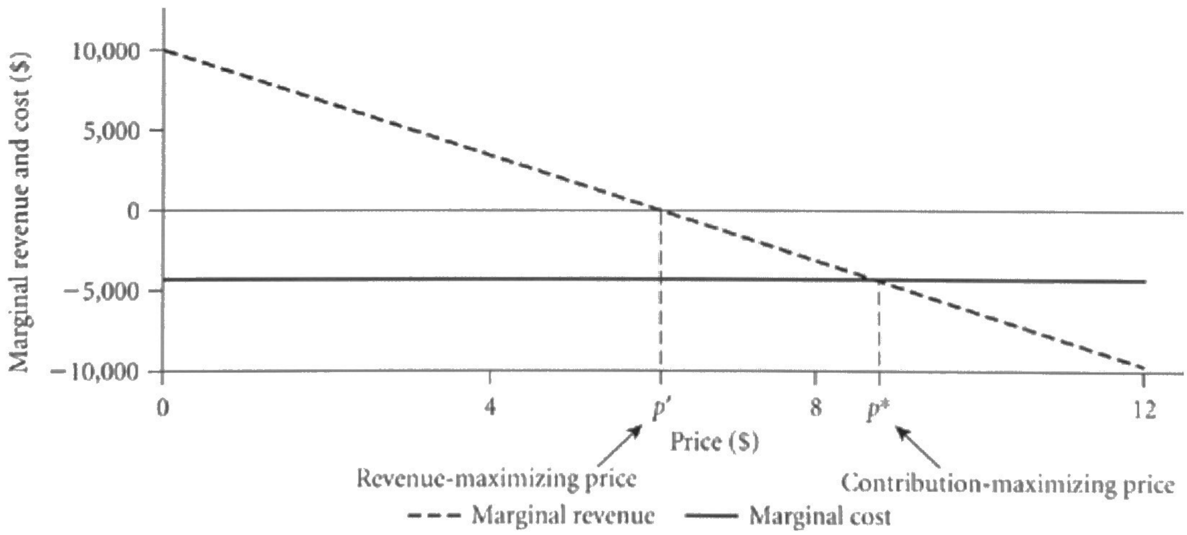

Figure 3.14 Revenue maximization and contribution maximization.

The decision that management needs to make in Example 3.12 is whether or not it is worth giving up a total contribution of $5,000 per month to "buy" an additional 2,000 units of demand. Since the revenue-maximizing price is lower than the contribution-maximizing price, there is no guarantee that the revenue-maximizing price will provide a reasonable margin—or even a positive margin—if incremental cost is greater than zero. For this reason, it is dangerous to maximize total revenue without including a constraint that ensures that the resulting price is greater than incremental cost.

3.4.5 Weighted Combinations of Revenue and Contribution

In some cases, a company might wish to maximize a weighted combination of total contribution and revenue. Some pricing and revenue optimization systems use "slider bars" to determine how much to weight revenue relative to contribution. The most common approach to combining revenue contribution is to use a weighting parameter α, with $0 \leq \alpha \leq 1$, resulting in a weighted objective function $Z(p)$ given by

$$Z(p) = \alpha(p - c)\, d(p) + (1 - \alpha) p\, d(p) \tag{3.25}$$

For $\alpha = 1$, $Z(p)$ equals total contribution; for $\alpha = 0$, $Z(p)$ equals revenue. Values of α between 0 and 1 will maximize a weighted combination of the two, with higher values of α resulting in a higher weighting for contribution relative to total revenue.

Applying some algebra to Equation 3.25 gives

$$Z(p) = (p - \alpha c)\, d(p) \tag{3.26}$$

which shows that maximizing a weighted combination of the revenue and total contribution is the same as maximizing contribution with a discounted cost. For example, a value of $\alpha = 0.5$ implies equal weights for total revenue and total contribution. Equation 3.26 shows that the price that maximizes this weighted objective function can be found by maximizing total contribution with cost reduced by 50%. Equation 3.25 allows us to state a general principle.

The price that maximizes a weighted combination of revenue and contribution is greater than or equal to the revenue-maximizing price and less than or equal to *the contribution-maximizing price.*

In the absence of other constraints, we would only be interested in prices greater than the revenue-maximizing price and less than the contribution-maximizing price. In other words, there is no reason for an unconstrained seller to consider pricing outside of this range.

3.5 SUMMARY AND EXTENSIONS

1. The core problem in PRO can be formulated as a constrained optimization problem where the objective function is to maximize total contribution. The constraints are the result of either business rules (e.g., the desire to maintain a minimum market share in a certain market) or constraints on capacity or inventory.
2. A key input into any PRO problem is the price-response function that relates price to demand. The price-sensitivity function is typically nonnegative, continuous, and downward sloping.
3. Two common key measures of price sensitivity are the slope and elasticity of the price-response function, where the slope is defined as the derivative of the price-response function and elasticity is the (approximate) percentage change in demand that would result from a 1% change in price. Both slope and elasticity are *local* properties of the price-response function, in that they can be used to estimate the effects of small changes in price but not large changes.
4. In many cases, price-response functions can be considered as the measure of the number of people whose maximum willingness to pay (or reservation price) is greater than a certain price. In this case, a price-response function corresponds to a particular distribution of willingness to pay across a population. For example, a linear price-response function corresponds to a uniform distribution on willingness to pay.