PART III

COMMON ELEMENTS

Chapter 7

CUSTOMER-BEHAVIOR AND MARKET-RESPONSE MODELS

This chapter reviews the basic theory of consumer choice, aggregate demand, and the operational, market-response models that are used in both quantity- and price-based revenue management. Because demand results from many individuals making choice decisions—choices to buy one firm's products over another, to wait or not to buy at all, to buy more or fewer units—we begin by looking at models of individual-choice behavior. When added up, these individual purchase decisions determine aggregate demand, so we next discuss aggregate-demand functions and their properties. Our treatment of the theory is somewhat abbreviated, aimed more at developing an intuitive and practical understanding of the concepts. The Notes and Sources section at the end of the chapter provides references that offer more extensive treatment of consumer behavior theory. Appendix E at the end of the book provides a basic reference on consumer theory, including utility theory, reservation prices, and risk preferences.

7.1 The Independent-Demand Model

Before delving into more complex models of demand, we first briefly review the *independent-demand model,* which is the basis of much of the material in Chapters 2 and 3 on quantity-based RM. This model is rather simple: it assumes that demand for each product is an independent stochastic process, not influenced by the firm's availability controls. Further, as we have seen in Section 2.2, static models of quantity-based RM also assume that the demand for products arrives in a specified order over the booking period, with demand for the lower-priced products appearing first. Thus, the independent model does not endogenize customer behavior, neither choice behavior nor purchase-timing behavior.

While it is easy to criticize the simplistic nature of this model, one can make a few theoretical and practical arguments in support of it. As discussed in Chapter 2, in standard quantity-based RM practice the customer is faced with a menu of possible products differentiated by prices and restrictions. As a result, if the firm offers n products, customers are approximately segmented into n separated populations (one for each product) according to their preference for the different product restrictions and prices. If customers are sufficiently well segmented by the restrictions (in the sense that most of the customers who are eligible to purchase one product are not eligible to buy another), then the independent-demand assumption is not unreasonable. However, this argument is admittedly weakened by the fact that (at least in the airline case) most restrictions are progressively relaxed as the fares get higher. So a customer who is eligible for one fare class is normally eligible for all classes with higher fares. We must then assume that customers are unwilling to purchase these higher fares.

Second, the independent-demand model is reasonable if the market is competitive and products are commodities—defined as products in which the identity of the supplier is of little importance to customers. In such cases, firms are price takers and can control only the quantity they sell; customers, in turn, base their choices only on price and are willing to buy from any firm offering the market price. (See Section 8.2.) Hence, if a given commodity product is not available at one firm—in particular, if its availability is closed by RM system controls—then demand for that product effectively disappears because customers will purchase the product from a competitor rather than switching to alternative products.

Third, the model reflects the current airline and hotel industry practice of separating pricing and capacity-control decisions, reflecting the different scope of the two in these industries—pricing decisions are made infrequently, while capacity-control is done in real-time; prices are set at a market level that includes a large number of flight departures (for airlines) and for an entire season (for a hotel) while capacity control is exerted on individual flights and dates. The implicit assumption in traditional quantity-based RM is that when prices for the products change, the change in the demand being observed will influence the forecasts and this changed forecasts, together with the new price of each product, will lead to changes in the capacity controls on each flight. This sort of quasi-static view of the price-demand relationship lies at the heart of current RM practice, and indeed the success of traditional RM methodology points to the practical utility of the overall approach.

Finally, the independent-demand model considerably simplifies the RM forecasting and optimization tasks. Forecasting can use historical demand data in standard time series forecasting methods, and we can solve stochastic optimization models based on the independent-demand model (at least approximately) relatively efficiently.

Yet despite these arguments in support of the independent-demand model, the fact that it ignores consumer behavior is conceptually unsatisfying and, more important, limits the full potential of RM methods. To counter its simplifications, a number of ad-hoc methods, such as the sell-up model discussed in Section 2.6, have been proposed. The discrete-choice model of Section 2.6 is a more recent alternative that overcomes the limitations of the independent-demand model. This latter model has more in common with the customer-choice-behavior view of demand that is the focus of this chapter.

7.2 Models of Individual Customer Choice

We next look at the basic approaches for modeling individual customer purchase decisions. In Chapter 9, we discuss methods for estimating the parameters of these models.

7.2.1 Reservation-Price Models

The simplest practical models of customer choice directly model customers' reservation prices for particular items. Each customer is assumed to follow a simple decision rule: if his reservation price (or valuation) v equals or exceeds the offered price p, the customer purchases the product; otherwise, he will not purchase the product. Moreover, he buys at most one unit of the product.

A customer's reservation price is specific to each individual and typically is private information unknown to the firm. However, based on management judgment, historical observed purchase behavior or other observable characteristics of the individual (such as place, time, and channel of purchase), the seller can attempt to model the distribution of the reservation prices across a population of customers and estimate at least the parameters of the distribution. This leads to a problem of finding a distribution $F(\cdot)$ such that the probability that a customer's reservation price is below p is given by $F(p) = P(v \leq p)$.

Often, however, the distribution of reservation prices is modeled indirectly by assuming an aggregate-demand function, as we discuss in Section 7.3 below. Hence, we postpone further discussion of reservation-price modeling until that point.

7.2.2 Random-Utility Models

Random-utility models are based on a probabilistic model of individual customer utility. (See Appendix E for a formal discussion of utility theory.) They are useful for several reasons. First, probabilistic models can be used to represent heterogeneity of preference among a population of customers. Second, they can model uncertainty in choice outcomes due to the inability of the firm to observe all the relevant variables affecting a given customer's choice (other alternatives, their prices, the customer's wealth, and so on). Third, they can model situations where customers exhibit *variety-seeking behavior* and deliberately alter their choices over time (movie or meal choice, for example). Finally, probabilistic choice can model customers whose behavior is inherently unpredictable—that is, customers who behave in a way that is inconsistent with well-defined preferences and at best, exhibit only some probabilistic tendency to prefer one alternative to another. Luce [349] developed a model of this type of random-choice behavior based purely on a set of axioms on choice probabilities, analogous to the axioms used to define classical deterministic utility functions. (See Appendix E.)[1]

For all these reasons, it is often reasonable to assume that a firm has only probabilistic information on the utility function of any given customer, and this can be modeled by assuming that customers' utilities for alternatives are themselves random variables. Specifically, let the n alternatives be denoted $j = 1, \ldots, n$. A customer has a utility for alternative j, denoted U_j. Without loss of generality we can decompose this utility into two parts, a *representative* component u_j that is deterministic and a mean-zero *random* component, ξ_j. Therefore,

$$U_j = u_j + \xi_j, \tag{7.1}$$

and the probability that an individual selects alternative j from a subset S of alternatives is given by[2]

$$P_j(S) = P(U_j \geq \max\{U_i : i \in S\}). \tag{7.2}$$

In other words, the probability that j has the highest utility among all the alternatives in the set S.

The representative component u_j is often modeled as a function of various observable attributes of alternative j. A common assumption is

[1]The distinction between models based on randomized preferences and those based on random-choice behavior is important primarily to behavioral theorists. A seminal work in this area is Block and Marschak [79]. However, for most RM problems what matters most is the demand process produced by a given model.

[2]That customers choose based on maximizing utility is itself an assumption. See Appendix E for a discussion of utility maximization as a model of customer choice.

the *linear-in-attributes* model

$$u_j = \boldsymbol{\beta}^\top \mathbf{x}_j, \tag{7.3}$$

where β is a vector of parameters and $\mathbf{x}_j$ is a vector of attribute values for alternative j, which could include factors such as price, measures of quality and indicator variables for product features. Variables describing characteristics of the customer (segment variables) can also be included in $\mathbf{x}_j$.[3]

This formulation defines a general class of random-utility models, which vary according to the assumptions on the joint distribution of the utilities $U_1, \ldots, U_n$. Random-utility models are no more restrictive in terms of modeling behavior than are classical utility models; essentially, all we need assume is that customers have well-defined preferences so that utility maximization is an accurate model of their choice behavior. (Theorem E.3 in Appendix E.) However, as a practical matter, certain assumptions on the random utilities lead to much simpler models than others. We look at a few of these special cases next.

7.2.2.1 Binary Probit

If there are only two alternatives to choose from (such as buying or not buying a product) and the error terms $\xi_j, j = 1, 2$, are independent, normally distributed random variables with mean zero and identical variances σ^2, then the probability that alternative 1 is chosen is given by

$$P(\xi_2 - \xi_1 \leq u_1 - u_2) = \Phi(\frac{u_1 - u_2}{\sqrt{2}\sigma}), \tag{7.4}$$

where $\Phi(\cdot)$ denotes the standard normal distribution. This model is known as the *binary-probit* model. While the normal distribution is an appealing model of disturbances in utility (it can be viewed as the sum of a large number of random disturbances), the resulting probabilities do not have a closed-form solution. This has led researchers to seek other, more analytically tractable, models.

7.2.2.2 Binary Logit

The binary-logit model applies also to a situation with exactly two choices, similar to the binary-probit case, but is simpler to analyze. The assumption made here is that the error term $\xi = \xi_1 - \xi_2$ has a *logistic*

[3]Specifically, the utility can also depend on observable customer characteristics, so for customer i the utility of alternative j is u_{ij}. For simplicity, we ignore customer-specific characteristics here, but they can be incorporated into all the models that follow.

distribution—that is,

$$F(x) = \frac{1}{1 + e^{-\frac{x}{\mu}}},$$

where $\mu > 0$ is a scale parameter and $-\infty < x < \infty$. Here ξ has a mean zero and variance $\frac{\mu^2\pi^2}{3}$. The logistic distribution provides a good approximation to the normal distribution, though it has "fatter tails." The probability that alternative 1 is chosen is given by

$$P(\xi_2 - \xi_1 \leq u_1 - u_2) = \frac{e^{\frac{u_1}{\mu}}}{e^{\frac{u_1}{\mu}} + e^{\frac{u_2}{\mu}}}. \qquad (7.5)$$

7.2.2.3 Multinomial Logit

The multinomial-logit model (MNL) is a generalization of the binary-logit model to n alternatives. It is derived by assuming that the ξ_j are i.i.d. random variables with a Gumbel (or double-exponential) distribution with cumulative density function

$$F(x) = P(\xi_j \leq x) = e^{-e^{-(\frac{x}{\mu}+\gamma)}},$$

where γ is Euler's constant (= 0.5772...) and μ is a scale parameter. The mean and variance of ξ_j are

$$E[\xi_j] = 0, \qquad Var[\xi_j] = \frac{\mu^2\pi^2}{6}.$$

The Gumbel distribution has some useful analytical properties, the most important of which is that the distribution of the maximum of n independent Gumbel random variables with the same scale parameter μ is also a Gumbel random variable. If two random variables ξ_1 and ξ_2 are Gumbel distributed with mean 0 and scale parameter μ, then $\xi = \xi_1 - \xi_2$ has a logistic distribution with mean 0 and variance, $\frac{\mu^2\pi^2}{3}$, leading to the binary-logit model.

For the MNL model, the probability that an alternative j is chosen from a set $S \subseteq \mathcal{N} = \{1, 2, \ldots, n\}$ that contains j is given by

$$P_j(S) = \frac{e^{\frac{u_j}{\mu}}}{\sum_{i \in S} e^{\frac{u_i}{\mu}}}. \qquad (7.6)$$

If $\{u_j : j \in S\}$ has a unique maximum and $\mu \to 0$, then the variance of the $\xi_j, j = 1, \ldots, n$ tends to zero and the MNL reduces to a deterministic model—namely

$$\lim_{\mu \to 0} P_j(S) = \begin{cases} 1 & \text{if } u_j = \max_{i \in S}\{u_i\} \\ 0 & \text{otherwise.} \end{cases}$$

Conversely, if $\mu \to \infty$, then the variance of the $\xi_j, j = 1, \ldots, n$ tends to infinity and the systematic component of utility u_j becomes negligible. In this case,

$$\lim_{\mu \to \infty} P_j(S) = \frac{1}{|S|}, \quad j \in S,$$

which corresponds to a uniform random choice of the alternatives in *S*. Hence, the MNL can model behavior ranging from deterministic utility maximization to purely random choice.

The MNL has been widely used as a model of customer choice. However, it possesses a somewhat restrictive property known as the *independence from irrelevant alternatives* (IIA) property—namely, for any two sets $S \subseteq \mathcal{N}$, $T \subseteq \mathcal{N}$ and any two alternatives $i, j \in S \cap T$, the choice probabilities satisfy

$$\frac{P_i(S)}{P_j(S)} = \frac{P_i(T)}{P_j(T)}. \tag{7.7}$$

Equation (7.7) says that the relative likelihood of choosing i and j is independent of the choice set containing these alternatives. This property is not realistic, however, if the choice set contains alternatives that can be grouped such that alternatives within a group are more similar than alternatives outside the group because adding a new alternative reduces the probability of choosing similar alternatives more than dissimilar alternatives. A famous example illustrating this point is the "blue-bus/red-bus paradox," (Debreu [150]):

Example 7.1 An individual has to travel and can use one of two modes of transportation: a car or a bus. Suppose the individual selects them with equal probability. Let the set S = {car, bus}. Then

$$P_{\text{car}}(S) = P_{\text{bus}}(S) = \frac{1}{2}.$$

Suppose now that another bus is introduced that is identical to the current bus in all respects except color: one is blue and one is red. Let the set T denote {car, blue bus, red bus}. Then the MNL predicts

$$P_{\text{car}}(T) = P_{\text{blue bus}}(T) = P_{\text{red bus}}(T) = \frac{1}{3}.$$

However, as bus color is likely an irrelevant characteristic in this choice situation, it is more realistic to assume that the choice of bus or car is still equally likely, in which case we should have

$$P_{\text{car}}(T) = \frac{1}{2}$$
$$P_{\text{blue bus}}(T) = P_{\text{red bus}}(T) = \frac{1}{4}.$$

As a result of IIA, the MNL model must be used with caution. It should be restricted to choice sets that contain alternatives that are, in some sense, "equally dissimilar." Example 9.18 provides one empirical test for the IIA property.

Despite this deficiency, the MNL model is widely used in marketing. (See Guadagni and Little's [227] work on determining brand share in the presence of marketing variables such as advertising and promotion.) It has also seen considerable application in estimating travel demand. (See Ben-Akiva and Lerman [48].) The popularity of MNL stems from the fact that it is analytically tractable, relatively accurate (if applied correctly), and can be estimated easily using standard statistical techniques. (See Example 9.6.)

Variations of the MNL have been introduced to avoid the IIA problem, the most prevalent of which is the nested MNL [49]. Our next section looks at some generalizations of the MNL that avoid the IIA property.

7.2.3 Customer Heterogeneity and Segmentation

RM often relies on the premise that different customers are willing to pay different amounts for a product. For example, demand functions arise from heterogeneity in the reservation prices of customers. In many situations, this level of modeling of heterogeneity is sufficient or is the only practical approach.

Yet a more accurate representation of demand is achievable if customers can be segmented into groups with similar preferences and price responses. This entails classifying customers into K segments, where each segment has its own choice model. If done properly, each of these segment-level models predicts the behavior of the segment better than a common choice model. In the extreme case, one could potentially define a different segment for each customer. However, a model of heterogeneity has to find the right balance between estimability and accuracy; each segment should not be so narrowly defined or so small as to make estimation impossible, yet it should be sufficiently small that customers within a segment have relatively homogeneous price and marketing variable responses. The aim is to maximize between-group variation but minimize within-group variation with respect to market responses. (Many of the techniques used to identify and segment customers are based on cluster analysis.) We next look at a few common approaches along these lines.

7.2.3.1 Finite-Mixture Logit Models

In the basic MNL model with linear-in-attribute utilities, the coefficients $\boldsymbol{\beta}$ in (7.3) are assumed to be the same for all customers. This may not be an appropriate assumption if there are different segments with different preferences. Moreover, as we've seen, the assumption leads to the IIA property, which may not be reasonable in certain contexts. If we can identify each customer as belonging to a segment, then it is an easy matter to simply fit a separate MNL model to the data from each segment. However, a more sophisticated modeling approach is needed if segment membership is not observable.

Assume that customers within each segment follow a MNL model with identical parameters and that customers have a certain probability of belonging to a segment (called a *latent segment),* which has to be estimated along with the MNL parameters for each segment. This results in the so-called *finite-mixture logit models.*

Assume that there are L latent segments and that the probability that a customer belongs to segment l is given by

$$q_l = \frac{e^{\nu_l}}{\sum_{i=1}^{L} e^{\nu_i}}, \quad l = 1, \ldots, L.$$

All customers in segment l are assumed to have utilities determined by an identical vector of coefficients $\boldsymbol{\beta}_l$. Then the probability of choosing alternative j in this finite-mixture logit model is given by

$$P_j(S) \;=\; \sum_{l=1}^{L} q_l \frac{e^{\boldsymbol{\beta}_l{}^{\top}\mathbf{x}^j}}{\sum_{i\in S} e^{\boldsymbol{\beta}_l{}^{\top}\mathbf{x}_i}}, \quad j \in S.$$

One then tries to estimate the coefficients of the model ($\boldsymbol{\beta}$ and q_l, $l = 1, \ldots, L$) using, for example, maximum-likelihood methods. This model often provides better estimates of choice behavior than the standard MNL model, at the expense of a more complicated estimation procedure.

7.2.3.2 Random-Coefficients Logit Models

Another approach to modeling heterogeneity is to assume that each customer has a distinct set of coefficients $\boldsymbol{\beta}$ that are drawn from a distribution—usually assumed normal for analytical convenience—over the population of potential customers. This leads to what is called the *random-coefficients logit model.* The coefficients may also be correlated, both within themselves as well as with the error term, though we focus here on the simpler case where the coefficients are mutually independent.

Here again the utility of alternative j is given, similar to the MNL model, as

$$U_j = \boldsymbol{\beta}^{\top}\mathbf{x}_j + \xi_j, \quad j = 1, \ldots, n.$$

However, $\boldsymbol{\beta}$ is now considered a vector of random coefficients, each element of which is assumed to be independent of both the other coefficients in $\boldsymbol{\beta}$ and the error term ξ_j. Furthermore, the components of $\boldsymbol{\beta}$

are assumed to be normally distributed with a vector of means **b** and a vector of standard deviations $\boldsymbol{\sigma}$. The components of the random vector $\boldsymbol{\beta}$ corresponding to characteristic m, denoted β_m, can be decomposed into

$$\beta_m = b_m + \sigma_m \zeta_m,$$

where ζ_m, $m = 1, \ldots, M$ is a collection of i.i.d. standard normal random variables.

It is convenient to express the utility as a systematic part and a mean-zero error term as before: To this end, define the composite random-error term

$$\nu_j \;=\; [\sum_{m=1}^{M} x_{mj}\sigma_m\zeta_m] + \xi_j, \quad j = 1, \ldots, n. \tag{7.8}$$

Then a customer's random utility is given by

$$U_j(\boldsymbol{\nu}) = \mathbf{b}^{\top}\mathbf{x}_j + \nu_j, \quad j = 1, \ldots, n, \tag{7.9}$$

where ν_j is given by (7.8). Hence, the key difference between the standard MNL and the random-coefficient logit is that the error terms ν_j are no longer independent across the alternatives (and somewhat less important, they are no longer Gumbel distributed). The following example illustrates the idea:

Example 7.2 Suppose that there are three alternatives $j = 1, 2, 3$ with two characteristics each $m = 1, 2$ and that the values of the characteristics are given as in Table 7.1.

If the parameter means are estimated as $b_1 = 1, b_2 = 1$, then the logit model would have a customer choosing one of the three products with an equal probability. In contrast, the random-coefficients logit model would have customers with a high preference for characteristic 1 (ζ_m high) consider alternatives 1 and 2 as closer substitutes than alternative 3. Customer preferences and product characteristics interact via (7.8).

Note also that the IIA property of standard logit is partially mitigated in this model. A customer with a high preference for characteristic $m = 1$ will choose alternative 2 with high probability if the choice set {2,3} is offered and will choose 1 or 2 with equal probability if the choice set {1,2, 3} is offered.

7.3 Models of Aggregate Demand

Even with transaction-level data, it is often easier to model and estimate aggregate demand rather than individual customer-choice decisions. Figure 7.1 illustrates how the heterogenous reservation prices of individual demand translate into a price versus quantity relationship for aggregate demand. Depending on the model, this aggregate demand

Table 7.1. Attribute weights x_m^j for attributes $m = 1, 2$ in alternative $j = 1, 2, 3$.

	Alternatives (j)		
m	1	2	3
1	200	200	-100
2	-100	-100	200

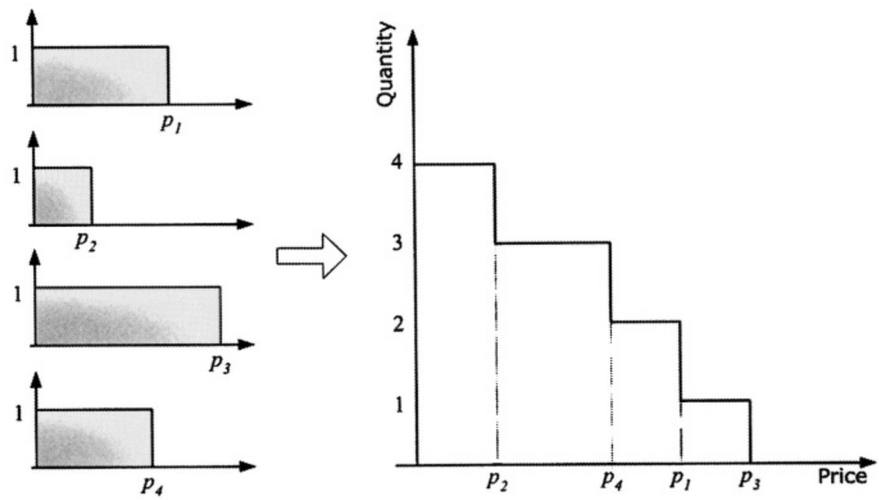

Figure 7.1. Individual demand with different reservation prices and the aggregate demand.

could be defined at the product, firm, or market level. If defined at the product or firm level, interactions with demand for other products (cross-elasticities) and dependence on historical demand or product attributes may have to be incorporated in its specification. In this section, we look at some commonly used aggregate-demand models.

7.3.1 Demand Functions and Their Properties

For the case of a single product, let p and $d(p)$ denote, respectively, the (scalar) price and the corresponding demand at that price. Also let Ω_p denote the set of feasible prices (the domain) of the demand function. For most demand functions of interest, $\Omega_p = [0, +\infty)$ but some functions (such as the linear-demand function) are not well defined for all nonnegative prices.

7.3.1.1 Regularity

It is often convenient to make the following regularity assumptions about the demand function:

ASSUMPTION 7.1 (REGULARITY: SCALAR CASE)
(i) The demand function is continuously differentiate on Ω_p.
(ii) The demand function is strictly decreasing, $d'(p) < 0$, *on* Ω_p.
(iii) The demand function is bounded above and below:

$$0 \leq d(p) < \infty, \quad \forall p \in \Omega_p.$$

(iv) The demand tends to zero for sufficiently high prices—namely,

$$\inf_{p \in \Omega_p} d(p) = 0.$$

(v) The revenue function $pd(p)$ *is finite for all* $p \in \Omega_p$ *and has a finite maximizer* $p^0 \in \Omega_p$ *that is interior to the set* Ω_p.

These are not restrictive assumptions in most cases and simply help avoid some technical complications in both analysis and numerical optimization. For example, consider a linear demand model (defined formally in Section 7.3.3.1)

$$d(p) = a - bp, \quad p \in \Omega_p = [0, a/b]. \tag{7.10}$$

This is trivially differentiable on Ω_p, is strictly decreasing if $b > 0$, is nonnegative and bounded for all $p \in \Omega_p$, tends to zero for $p \to a/b$ and the revenue $ap - bp^2$ and has a finite maximizer $p^0 = \frac{a}{2b}$.

7.3.1.2 Market-Share and Reservation-Price Distribution

It is sometimes convenient to express the demand function in the form

$$d(p) = N(1 - F(p)), \tag{7.11}$$

where $F(p)$ is a cumulative distribution function and N is interpreted as the *market size.* $1 - F(p)$ is then interpreted as the fraction of the market that is willing to buy at price p; equivalently, $F(p)$ is the distribution of reservation prices v in the customer population. The derivative of $F(p)$ is denoted $f(p) = \frac{\partial}{\partial p} F(p)$.

For example, consider again the linear-demand function (7.10). This can be written in the form (7.11) if we define $N = a$ and $F(p) = pb/a$. Since $F(\cdot)$ is the probability distribution of a customer's reservation price v, reservation prices are uniformly distributed in the linear-demand-function case.

7.3.1.3 Elasticity of Demand

The *price elasticity* of demand is the relative change in demand produced by a relative change in price. It is defined by

$$\epsilon(p) \equiv \frac{p}{d}\frac{\partial d}{\partial p} = \frac{\partial \ln(d)}{\partial \ln(p)}.$$

Note that elasticity is defined at a particular price p.

To illustrate, for the linear-demand function (7.10), $\frac{\partial d}{\partial p} = -b$, so the elasticity is

$$\frac{p}{d}\frac{\partial d}{\partial p} = -\frac{bp}{a - bp}.$$

Products can be categorized based on the magnitude of their elasticities. A product with $|\epsilon(p)| > 1$ is said to be *elastic,* while one with a elasticity value $|\epsilon(p)| < 1$ is said to be *inelastic.* If $|\epsilon(p)| = \infty$, demand for the product is said to be *perfectly elastic,* while if $|\epsilon(p)| = 0$, demand is said to be *perfectly inelastic.* Table 7.2 shows a sample of estimated elasticities for common consumer products. While many factors affect elasticity, these estimates give some sense of the relative magnitudes of elasticities.

7.3.1.4 Inverse Demand

The *inverse-demand function,* denoted $p(d)$, is the largest value of p which generates a demand equal to d—that is,

$$p(d) \equiv \max_{p \in \Omega_p} \{p : d(p) = d\}.$$

Given an inverse-demand function, one can view demand rather than price as the decision variable, since every choice of a demand d implies a unique choice of price $p(d)$. This is useful, as it is often easier analytically and computationally to work with demand rather than price as the decision variables in optimization problems.

The inverse may not be well-defined, however—for example, for values of d corresponding to points at which the demand function $d(p)$ has a jump discontinuity. Also, there may be not be a price p that produces any given value of demand d (for example, if demand remains bounded as p tends to zero yet d is large). Since not all values of d may be obtainable, we let Ω_d denote the set of achievable demand values. This set plays a role analogous to Ω_p for the demand function.

Under the regularity Assumption 7.1, the demand function is strictly decreasing and continuously differentiable on Ω_p, so the inverse-demand function is always well defined and continuously differentiable on the set $\Omega_d = \{x : x = d(p),\ p \in \Omega_p\}$. Indeed, under Assumption 7.1, the

Table 7.2. Estimated elasticities (absolute values) for common products.[a]

Product	$\vert\epsilon(p)\vert$
Inelastic	
Salt	0.1
Matches	0.1
Toothpicks	0.1
Airline travel, short-run	0.1
Gasoline, short-run	0.2
Gasoline, long-run	0.7
Residential natural gas, short-run	0.1
Residential natural gas, long-run	0.5
Coffee	0.25
Fish (cod) consumed at home	0.5
Tobacco products, short-run	0.45
Legal services, short-run	0.4
Physician services	0.6
Taxi, short-run	0.6
Automobiles, long-run	0.2
Approximately unit elasticity	
Movies	0.9
Housing, owner occupied, long-run	1.2
Shellfish, consumed at home	0.9
Oysters, consumed at home	1.1
Private education	1.1
Tires, short-run	0.9
Tires, long-run	1.2
Radio and television receivers	1.2
Elastic	
Restaurant meals	2.3
Foreign travel, long-run	4
Airline travel, long-run	2.4
Fresh green peas	2.8
Automobiles, short-run	1.2–1.5
Chevrolet automobiles	4
Fresh tomatoes	4.6

[a] *Source:* Reported by Gwartney and Stroup [232], collected from various econometric studies.

demand function is continuous, decreasing, and bounded and tends to zero for sufficiently high prices. One can also verify that the domain of the inverse-demand function is always an interval of the form $\Omega_d = [0, \bar{d}]$ for some upper bound $\bar{d}$.

Equation (7.11) expressed in terms of the reservation-price distribution, $F(p)$, the inverse-demand function is defined by

$$p(d) = F^{-1}(1 - d/N),$$

where $F^{-1}(\cdot)$ is the inverse of $F(\cdot)$.

To illustrate, the inverse of the linear-demand function (7.10) is

$$p(d) = \frac{1}{b}(a - d),$$

and the set of feasible demand rates is $\Omega_d = [0, a]$.

7.3.1.5 Revenue Function

The *revenue function,* denoted $r(d)$, is defined by

$$r(d) \equiv dp(d).$$

This is the revenue generated when using the price p and is of fundamental importance in dynamic-pricing problems. For example, the linear-demand function (7.10) has a revenue function

$$r(d) = \frac{d}{b}(a - d).$$

For most dynamic-pricing problems, we require that this revenue function be concave, as in the linear example above. This condition leads to well-behaved optimization problems.

7.3.1.6 Marginal Revenue

Another important quantity in pricing analysis is the rate of change of revenue with quantity—the *marginal revenue*—which is denoted $J(d)$. It is defined by

$$\begin{aligned} J(d) &\equiv \frac{\partial}{\partial d} r(d) \\ &= p(d) + dp'(d). \end{aligned} \qquad (7.12)$$

It is frequently useful to express this marginal revenue as a function of price rather than quantity. At the slight risk of confusion over notation, we replace d by $d(p)$ above and define the marginal revenue as a function of price by[4]

$$J(p) \equiv J(d(p)) = p + d(p)\frac{1}{d'(p)}. \qquad (7.13)$$

[4] By the inverse-function theorem, $p'(d) = 1/d'(p)$.

Note that $J(p)$ above is still the marginal revenue with respect to quantity—$\frac{\partial}{\partial d}r(d)$—but expressed as function of price rather than quantity; in particular, it is not the marginal revenue with respect to price.[5]

Expressing marginal revenue in terms of the reservation-price distribution $F(p)$, we have that

$$J(p) = p - \frac{1}{\rho(p)}, \tag{7.14}$$

where $\rho(p) \equiv f(p)/(1 - F(p))$ is the *hazard rate* of the distribution $F(p)$.[6] The marginal revenue function plays an important role in pricing problems. It is also central to the design of revenue-maximizing auctions, where it is referred to as the *virtual utility,* for reasons that are discussed in Chapter 6.

To illustrate, consider the marginal revenue of the linear-demand function of (7.10) as a function of d,

$$J(d) = \frac{\partial}{\partial d}\left[\frac{d}{b}(p - d)\right] = \frac{1}{b}(a - 2d).$$

Substituting $d(p) = a - bp$ for d above we obtain the marginal revenue as a function of price

$$J(p) = \frac{1}{b}(a - 2(a - bp)) = 2p - \frac{a}{b}.$$

It is frequently useful to make the following assumption about the marginal revenue:

ASSUMPTION 7.2 (MONOTONE MARGINAL REVENUE) *The marginal revenue $J(d)$ defined by (7.11) is strictly decreasing in the demand d. Equivalently, the marginal revenue $J(p)$ defined by (7.13) is strictly increasing in the price p.*

---

[5]The relationship between the marginal revenue with respect to price and quantity is as follows: since $r = pd$, then $\frac{\partial r}{\partial d} = d\frac{\partial p}{\partial d} + p$ and $\frac{\partial r}{\partial p} = p\frac{\partial d}{\partial p} + d$. Therefore, $\frac{\partial r}{\partial p} = \frac{\partial d}{\partial p}(p + d\frac{\partial p}{\partial d}) = \frac{\partial d}{\partial p}\frac{\partial r}{\partial d}$. (This also follows from the chain rule.)

[6]To see this, note that $\frac{\partial}{\partial x}F^{-1}(x) = 1/f(x)$, so

$$\begin{aligned} J(p) &= p + d(p)\frac{1}{d'(p)} \\ &= p - d(p)\frac{1}{Nf(p))} \\ &= p - \frac{1 - F(p)}{f(p)}, \end{aligned}$$

where the first equality follows from (7.13) and the next two from (7.11).

Note that this condition guarantees that the revenue function $r(d)$ is a concave function of the demand d, which again is a useful property in optimization models because it guarantees that first-order conditions are sufficient for determining an optimal price. This property is satisfied, for example, by the linear-demand function.

Slightly weaker conditions than those of Assumption 7.2 will also ensure that pricing-optimization problems are well behaved. In particular, if the revenue function is strictly unimodal,[7] this is often sufficient to ensure that there is a unique optimal price. (This is true for simple unconstrained pricing problems, for example.) Recall that $f(p)$ denotes the reservation-price density (derivative of $F(p)$) and $\rho(p) = f(p)/(1-F(p))$ denotes the hazard rate. Then the following sufficient conditions on the reservation-price distribution ensure strict unimodality of the revenue function (see Ziya et al. [591]):

PROPOSITION 7.1 *Suppose that the reservation-price distribution $F(p)$ is twice differentiable and strictly increasing on its domain $\Omega_p = [p_1, p_2]$ ($F(p_1) = 0$ and $F(p_2) = 1$). Suppose further that $F(\cdot)$ satisfies any one of the following conditions:*
(i) $2\rho(p) > -\frac{f'(p)}{f(p)}$ *for all* $p \in \Omega_p$.
(ii) $\frac{2}{p} > -\frac{f'(p)}{f(p)}$ *for all* $p \in \Omega_p$.
(iii) $\rho(p) + \frac{1}{p} > -\frac{f'(p)}{f(p)}$ *for all* $p \in \Omega_p$

Then the revenue functions $r(p) = pd(p) = pN(1 - F(p))$ is strictly unimodal on Ω_p (equivalently, the revenue function $r(d) = p(d)d$ is strictly unimodal on $\Omega_d = [N(1 - F(p_1)), N(1 - F(p_2))]$.

Ziya et al. [591] show there are demand functions that satisfy one condition but not the others, so the three conditions are distinct.

Another desirable property of the marginal revenue function is that it spans the range $[0, +\infty)$ as p ranges over Ω_p (equivalently, d ranges over Ω_d). This is because in optimization problems, the first-order conditions typically involve equating marginal revenue to a nonnegative value (such as a cost or a Lagrange multiplier). If the marginal revenue spans the range $[0, +\infty)$, then the solutions of the first-order conditions are always in Ω_p (or Ω_d), and therefore, the explicit price (or demand) constraints can be safely ignored. We formalize this property in the following assumption:

[7] A function $f(x)$ defined on the domain $[a, b]$ is said to be a *unimodal function* if there exists an $x^* \in [a, b]$ such that $f(x)$ is strictly increasing on $[a, x^*]$ and $f(x)$ is strictly decreasing on $[x^*, b]$.

ASSUMPTION 7.3 *The range of the marginal revenue defined by (7.11) and (7.13) spans* $[0, +\infty)$. *That is, for every* $x \in [0, +\infty]$, $\exists d \in \Omega_d$ *such that* $J(d) = x$*; equivalently,* $\exists p \in \Omega_p$ *such that* $J(p) = x$.

Note that the linear-demand function does not satisfy this condition because the marginal revenue is $J(d) = (1/b)(a - 2d)$ and $\Omega_d = [0, a]$, so the marginal revenue ranges over $[-a/b, a/b]$. Other common demand functions, however, do satisfy this assumption, as described below.

7.3.1.7 Revenue-Maximizing Price

Under Assumption 7.2, the revenue is maximized at the point where the marginal revenue becomes zero. Assumption 7.1, part (v), requires that the maximizer is an interior point of the domain Ω_p, in which case the *revenue-maximizing price* p^0 is determined by the first-order condition

$$J(p^0) = 0.$$

Similarly, the revenue-maximizing demand, denoted d^0, is defined by

$$J(d^0) = 0.$$

They are related by

$$d^0 = d(p^0).$$

For example, for the linear-demand function we have $J(p) = 2p - a/b$ so $p^0 = \frac{a}{2b}$, an interior point of the set $\Omega_p = [0, a/b]$. The revenue-maximizing demand is $d^0 = a/2$.

Note from (7.13) that since $J(p) = p(1 + \frac{d}{p}\frac{\partial p}{\partial d})$, $\frac{\partial d}{\partial p} < 0$ (from Assumption 7.1, part (ii)), and $\frac{\partial d}{\partial p}\frac{p}{d} = \epsilon(p)$ is the price elasticity, we have

$$J(p) = p\left(1 - \frac{1}{|\epsilon(p)|}\right). \tag{7.15}$$

Thus, marginal revenue is increasing if demand is elastic at p (that is, if $|\epsilon(p)| > 1$), and marginal revenue is decreasing if demand is inelastic at p (that is, if $|\epsilon(p)| < 1$). At the critical value $|\epsilon(p^0)| = 1$, marginal revenue is zero and revenues are maximized.

If $J(p)$ is not monotone but one of the conditions of Proposition 7.1 is satisfied, then p^0 is a price such that $r(p^0)$ is increasing for $p < p^0$ and is decreasing for $p > p^0$; moreover, $p^0 = \inf\{p : |\epsilon(p)| \geq 1\}$.

Figure 7.2 illustrates the idea. Here, the revenue function $r(d)$ for the linear-demand function is plotted above, while the marginal-revenue function $J(d)$ is plotted below. Moving to the right corresponds to increasing the demand d and decreasing the price p. The inelastic-demand region is to the right of p^0, and the elastic region is to the left of p^0.

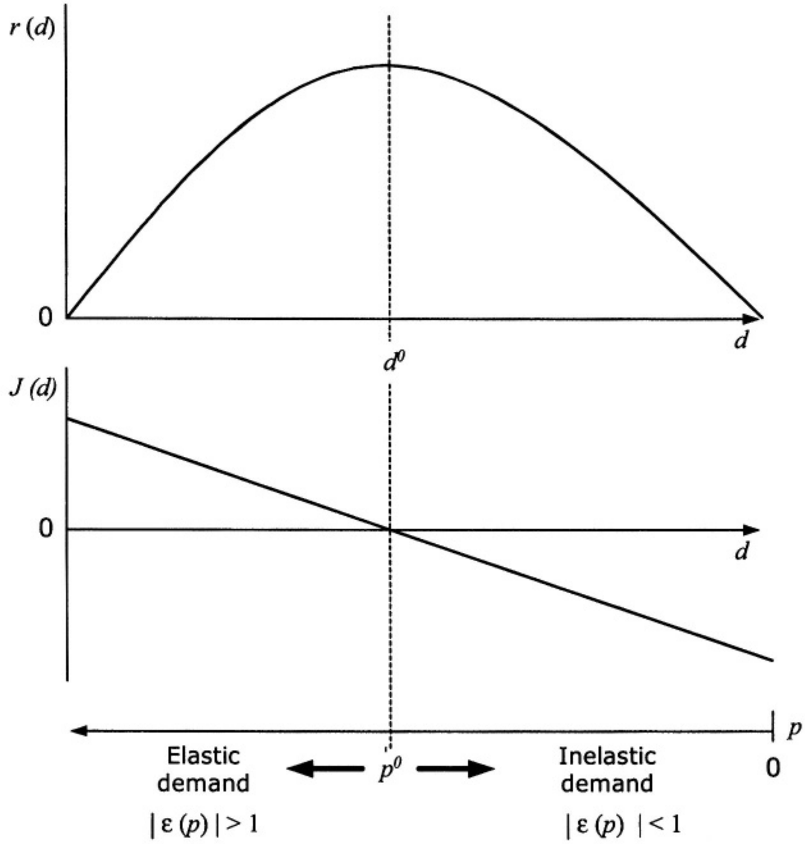

Figure 7.2. **Revenue and marginal-revenue curves.**

Starting at the far right with a price of zero, the demand is very inelastic; large relative changes in price (for example, doubling the price from δ to 2δ) result in small relative changes in demand. As a result, raising the price increases revenues. To the left, at very high price levels, relatively small decreases in price result in large increases in demand. Consequently, decreasing price improves revenues. The optimal price p^0 is the boundary of these two regions.

If there is a cost for providing the product—either a direct cost or opportunity cost—it is always optimal to price in the elastic region. To see this, let $c(d)$ denote the cost, so that $r(d) - c(d)$ is the firm's profit. Then the optimal price will occur at a point where $J(d) = r'(d) = c'(d)$. Assuming cost is strictly increasing in quantity, $c'(d) > 0$, the optimal price will be at a point where marginal revenue is positive—in the elastic region. Thus, it is almost never optimal to price in the inelastic region.[8]

7.3.2 Multiproduct-Demand Functions

In the case where there are $n > 1$ products, let p_j denote the price of product j and $\mathbf{p} = (p_1, \ldots, p_n)$ denote the vector of all n prices. The demand for product j as a function of $\mathbf{p}$ is denoted $d_j(\mathbf{p})$, and $\mathbf{d}(\mathbf{p}) = (d_1(\mathbf{p}), \ldots, d_n(\mathbf{p}))$ denotes the vector of demands for all n products. Again, $\mathbf{\Omega}_p$ will denote the domain of the demand function. We also use the notation $\mathbf{p}_{-j} = (p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_n)$ to denote all prices other than p_j.

Paralleling the single-product case, the following regularity assumptions for the multiproduct-demand function help ensure the resulting optimization models are well behaved:

ASSUMPTION 7.4 (REGULARITY: n-PRODUCT CASE) *For* $j = 1, \ldots, n$:
(i) $d_j(\mathbf{p})$ *demand is strictly decreasing in* p_j *for all* $\mathbf{p} \in \mathbf{\Omega}_p$.
(ii) The demand function is continuously differentiable on $\mathbf{\Omega}_p$.
(iii)The demand function is bounded above and below: $0 \le d_j(\mathbf{p}) < +\infty, \quad \forall \mathbf{p} \in \mathbf{\Omega}_p$.
(iv) The demand function tends to zero in its own price for sufficiently high prices—that is, for all $\mathbf{p}_{-j}$, $\inf_{p_j \in \mathbf{\Omega}_p} d_j(p_j, \mathbf{p}_{-j}) = 0$.
(v) The revenue function $\mathbf{p}^\top \mathbf{d}(\mathbf{p})$ *is bounded for all* $\mathbf{p} \in \mathbf{\Omega}_p$ *and has a finite maximizer* $\mathbf{p}^0$ *that is interior to* $\mathbf{\Omega}_p$.

As in the scalar case, we let $\mathbf{p}(\mathbf{d})$ denote the inverse-demand distribution; it gives the vector of prices that induces the vector of demands $\mathbf{d}$. In the multiproduct case, this inverse is more difficult to define generally, and in most cases we simply assume it exists. (For the common demand functions of Section 7.3.3, the inverse can be defined either explicitly or implicitly.) Likewise, we denote by $\mathbf{\Omega}_d$ the domain of the inverse-demand function, the set of achievable demand vectors $\mathbf{d}$.

The revenue function is defined by

$$r(\mathbf{d}) = \mathbf{d}^\top \mathbf{p}(\mathbf{d}),$$

which again represents the total revenue generated from using the vector of demands $\mathbf{d}$—or equivalently, the vector of prices $\mathbf{p}(\mathbf{d})$. Paralleling

[8]The only exception is if the firm *benefits* from disposing of products—that is, if it has a negative cost. For example, this could occur if there is a holding cost incurred for keeping units rather than selling them. In such cases, it may be optimal to price in the inelastic region.

Assumptions 7.2 and 7.3, in the multiproduct case it is often convenient to make the following assumption:

ASSUMPTION 7.5 *The multiproduct revenue function satisfies*
(i) $r(\mathbf{d})$ *is jointly concave on* $\mathbf{\Omega}_d$.
(ii) For every $\mathbf{x} \in \Re^n_+$, *there exists a* $\mathbf{d} \in \mathbf{\Omega}_d$ *such that* $\nabla_{\mathbf{d}} r(\mathbf{d}) = \mathbf{x}$.

Again, these assumptions help simplify the resulting pricing optimization problems and, and though more difficult to check than in the single-product case, are satisfied by several common demand functions.

The *cross-price elasticity* of demand is the relative change in demand for product i produced by a relative change in the price of product j. It is defined by

$$\epsilon_{ij}(p) = \frac{p_j}{d_i}\frac{\partial d_i}{\partial p_j} = \frac{\partial \ln(d_i)}{\partial \ln(p_j)}.$$

If the sign of the elasticity is positive, then products i and j are said to be *substitutes*; if the sign is negative, the products are said to be *complements.* Intuitively, substitutes are products that represent distinct alternatives filling the same basic need (such as Coke and Pepsi), whereas complements are products that are consumed in combination to meet the same basic need (such as hamburgers and buns).

7.3.3 Common Demand Functions

The demand function of a product can depend on variables other than its price (such as product attributes or, marketing variables such as advertising, brand name, competitor's prices and past market share), and modeling demand as a function of all relevant variables makes a model more realistic and accurate. The variables can either be current or lagged, when past-period variables affect demand. Here we focus on demand functions that depend only on current prices. A few other market-response functions that include nonprice variables are discussed in Section 9.6.4.

Table 7.3 summarizes the most common demand functions and their properties, and Figure 7.3 shows graphs of a few of these. All these functions satisfy the regularity conditions in Assumptions 7.1 and 7.4, the exception being the constant-elasticity-demand function, which does not satisfy part (v) of either assumption as explained below.

Table 7.3. Common demand functions.

	$d(p)$	$p(d)$	$r(d)$	$J(d)$	$\|\epsilon(p)\|$	p^0
Linear	$a - bp$	$\frac{1}{b}(a-d)$	$\frac{d}{b}(a-d)$	$\frac{1}{b}(a-2d)$	$\frac{pb}{a-bp}$	$\frac{a}{2b}$
Log-linear (exponential)	ae^{-bp}	$\frac{1}{b}\ln(\frac{a}{d})$	$\frac{d}{b}\ln(\frac{a}{d})$	$\frac{1}{b}(\ln(\frac{a}{d})-1)$	pbe^{-1}	$\frac{e}{b}$
Constant elasticity	ap^{-b}	$(\frac{a}{d})^{1/b}$	$a^{1/b}d^{1-1/b}$	$(1-\frac{1}{b})(\frac{a}{d})^{1/b}$	b	$\begin{cases} 0 & b>1 \\ +\infty & b<1 \\ \text{all } p \geq 0 & b=1 \end{cases}$
Logit	$\frac{e^{-bp}}{1+e^{-bp}}$	$\frac{1}{b}\ln(\frac{1}{d}-1)$	$\frac{d}{b}\ln(\frac{1}{d}-1)$	$\frac{1}{b}\left(\ln(\frac{1}{d}-1)-\frac{1}{d-1}\right)$	$\frac{p}{b(1+e^{-bp})}$	No closed-form formula

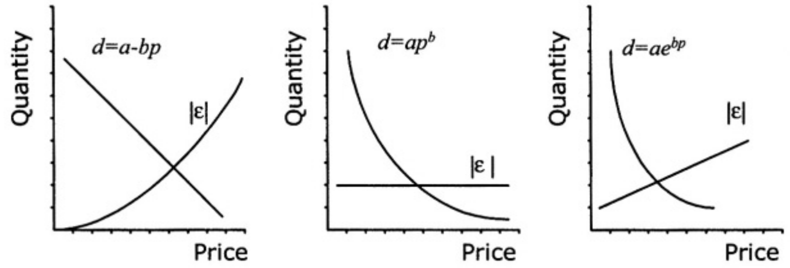

Figure 7.3. Some common aggregate demand (price-response) functions: (i) linear-demand function (ii) constant-elasticity-demand function (iii) exponential (log-linear) demand function.

7.3.3.1 Linear Demand

We have already seen the case of a linear-demand function in the scalar case. To summarize, it is

$$d(p) = a - bp,$$

where $a \geq 0$ and $b \geq 0$ are scalar parameters. The inverse-demand function is

$$p(d) = \frac{1}{b}(a - d).$$

The linear model is popular because of its simple functional form. It is also easy to estimate from data using linear-regression techniques. However, it produces negative demand values when $p > a/b$, which can cause numerical difficulties when solving optimization problems. Moreover, as mentioned, it does not satisfy Assumption 7.3. Hence, one must typically retain the price constraint set $\Omega_p = [0, a/b]$ when using the linear model in optimization problems.

In the multiproduct case, the linear model is

$$\mathbf{d}(\mathbf{p}) = \mathbf{a} + \mathbf{B}\mathbf{p},$$

where $\mathbf{a} = (a_1, \ldots, a_n)$ is vector of coefficients and $\mathbf{B} = [b_{ij}]$ is a matrix of price sensitivity coefficients with $b_{ii} \leq 0$ for all i and the sign of b_{ij}, $i \neq j$ depending on whether the products are complements $(b_{ij} < 0)$ or substitutes $(b_{ij} > 0)$. If $\mathbf{B}$ is nonsingular, then the inverse-demand function exists and is given by

$$\mathbf{p}(\mathbf{d}) = \mathbf{B}^{-1}(\mathbf{d} - \mathbf{a}).$$

One sufficient condition for $\mathbf{B}^{-1}$ to exist is that the *row* coefficients satisfy[9]

$$b_{ii} < 0 \text{ and } |b_{ii}| > \sum_{j \neq i} |b_{ij}|, \quad i = 1, \ldots, n. \tag{7.16}$$

Roughly, this says that demand for each product i is more sensitive to a change in its own price than it is to a simultaneous change in the prices of all other products. An alternative sufficient condition for $\mathbf{B}^{-1}$ to exist is that the *column* coefficients satisfy

$$b_{jj} < 0 \text{ and } |b_{jj}| > \sum_{i \neq j} |b_{ij}|, \quad j = 1, \ldots, n. \tag{7.17}$$

Equation (7.17) says that changes in the price of product j impacts the demand for product j more than it does the total demand for all other products combined. In the case of substitutes $(b_{ij} > 0, i \neq j)$, this is equivalent to saying there is an aggregate market expansion or contraction effect when prices change (for example, the total market demand strictly decreases when the price of product j increases, and demand for product j is not simply reallocated one for one to substitute products).

7.3.3.2 Log-Linear (Exponential) Demand

The log-linear—or exponential—demand function in the scalar case is defined by

$$d(p) = e^{a-bp},$$

where $a \geq 0$ and $b \geq 0$ are scalar parameters. This function is defined for all nonnegative prices, so $\Omega_p = [0, +\infty)$. The inverse-demand function is

$$p(d) = \frac{1}{b}(a - \ln(d)).$$

The log-linear-demand function is popular in econometric studies and has several desirable theoretical and practical properties. First, unlike the linear model, demand is always nonnegative so one can treat price (or quantity) as unconstrained in optimization problems. Second, by taking the log of demand, we recover a linear form, so it is also well suited to estimation using linear regression. However, demand values of zero are not defined when taking logarithms, which is problematic in settings where sales are infrequent.

The multidimensional log-linear form is

$$d_j(\mathbf{p}) = e^{a_j + \mathbf{B}_j^\top \mathbf{p}}, \quad j = 1, \ldots, n,$$

[9] As noted by Maglaras and Meissner [354] from conditions in Horn and Johnson [258].

where a_j is a scalar coefficient and $\mathbf{B}_j = (b_{j1}, \ldots, b_{jn})$ is a vector of price-sensitivity coefficients. Letting $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{B} = [b_{ij}]$, as in the linear model, and taking the logarithm, we have

$$\ln(\mathbf{d}(\mathbf{p})) = (\ln(d_1(\mathbf{p})), \ldots, \ln(d_n(\mathbf{p}))) = \mathbf{a} + \mathbf{Bp},$$

so again the log-linear model can be estimated easily from data using linear regression provided the data is not too sparse.

The inverse-demand function can be obtained as in the linear case if $\mathbf{B}$ is nonsingular, in which case

$$\mathbf{p}(\mathbf{d}) = \mathbf{B}^{-1}(\ln(\mathbf{d}) - \mathbf{a}),$$

and one can again use the sufficient conditions (7.16) or (7.17) to check that $\mathbf{B}^{-1}$ exists.

7.3.3.3 Constant-Elasticity Demand

The constant-elasticity-demand function in the single-product case is of the form

$$d(p) = ap^{-b},$$

where $a > 0$ and $b \geq 0$ are constants. The function is defined for all nonnegative p, so $\Omega_p = [0, +\infty)$. Since $\partial d/\partial p = -abp^{-(b+1)}$, the elasticity is

$$\epsilon(p) = \frac{p}{d}\frac{\partial d}{\partial p} = -b,$$

a constant for all values p (hence the name). The inverse-demand function is

$$p(d) = (\frac{a}{d})^{1/b}.$$

Note that because elasticity is constant, from (7.15) the marginal revenue will always be positive or will always be negative for all values of p (unless by chance $|\epsilon(p)| = 1$, in which case it is zero for all values of p). Thus, this function usually violates Assumption 7.1, part (iv), because either the marginal revenue is always positive so $p_0 = +\infty$ or the marginal revenue is always negative, so $p_0 = 0$, both extreme points of the set Ω_p (unless, again the elasticity is exactly one, in which case all values of p are revenue maximizing). From this standpoint, it is a somewhat ill behaved demand model in pricing-optimization problems, though in cases where revenue functions are combined with cost functions this behavior is less problematic.

The multiproduct constant elasticity model is

$$d_i(\mathbf{p}) = a_i p_1^{b_{i1}} p_2^{b_{i2}} \ldots p_n^{b_{in}}, \quad i = 1, \ldots, n,$$

where the matrix of coefficients $\mathbf{B} = [b_{ij}]$ defines the cross (and own) price elasticities among the products, since

$$\epsilon_{ij}(\mathbf{p}) = \frac{\partial d_i/d_i}{\partial p_j/p_j} = b_{ij}.$$

Note that the inverse-demand function $\mathbf{p(d)}$ exists if the matrix B is invertible, since $(\log(d_1(\mathbf{p})), \ldots, \log(d_n(\mathbf{p}))) = \mathbf{a} + \mathbf{B}p$ (here $\mathbf{a} = (a_1, \ldots, a_n)$) and $\log(\cdot)$ is a strictly increasing function.

7.3.3.4 Logit Demand

The logit demand function is based on the MNL model of Section 7.2.2.3. Recall that in the MNL the utility of each alternative $j = 1, \ldots, n$, is assumed to be of the form

$$U_j = u_j + \xi_j,$$

where u_j is the mean utility of choice j and ξ_j is an i.i.d., random-noise term with a Gumbel distribution with mean zero and scale parameter one. For the logit-demand function, we also include a no-purchase alternative (indexed by zero) with utility

$$U_0 = u_0 + \xi_0,$$

where ξ_0 is an independent Gumbel random variable with mean zero and scale parameter 1. Since utility is ordinal, without loss of generality we can assume $u_0 = 0$. The choice probabilities are then given by (7.6) with the no-purchase alternative having a value $e^{u_0} = 1$.

As mentioned, it is common to model u_j as a linear function of several known attributes including price. Assuming the representative component of utility u_j is linear in price and interpreting the choice probabilities as fractions of a population of customers of size N lead to the class of logit-demand functions.

For example, in the scalar case, we assume $u_1 = -bp$, and this gives rise to a demand function of the form

$$d(p) = N\frac{e^{-bp}}{1+e^{-bp}},$$

where N is the market size, $1 - F(p) = \frac{e^{-bp}}{1+e^{-bp}}$ is the probability that a customer buys at price p, and b is a coefficient of the price sensitivity. The function is defined for all nonnegative p, so $\Omega_p = [0, +\infty)$. There is no closed-form expression for the inverse-demand function, but it is easy to see that $d(p)$ is strictly decreasing in p, so the inverse exists and is well defined.

In the multiple-product case, the demand function is given by

$$d_j(\mathbf{p}) = N \frac{e^{-b_j p_j}}{1 + \sum_{i=1}^{n} e^{-b_i p_i}}, \quad j = 1, \ldots, n,$$

where again $\mathbf{b} = (b_1, \ldots, b_n)$ is a vector of coefficients and

$$P_j(\mathbf{p}) = \frac{e^{-b_j p_j}}{1 + \sum_{i=1}^{n} e^{-b_i p_i}}$$

is the MNL probability that a customer chooses product j as a function of the vector of prices $\mathbf{p}$.

One potential problem with the MNL demand model is that it inherits the IIA property (7.7). This causes problems if groups of products share attributes that strongly affect the choice outcome. To illustrate what can go wrong, consider the cross-price elasticity of alternative i with respect to the price of alternative j, $\epsilon_{ij}(\mathbf{p})$. This is given by

$$\begin{aligned} \epsilon_{ij}(\mathbf{p}) &= \frac{\partial \ln d_i(p)}{\partial \ln p_j} \\ &= -p_j b_j \frac{e^{-b_j p_j}}{1 + \sum_{k=1}^{n} e^{-b_k p_k}}. \end{aligned} \tag{7.18}$$

Notice that this cross-price elasticity is not dependent on i, and therefore cross-elasticity is the same for all alternatives i other than j.

The implications of this constant cross-price elasticity can be illustrated by an example of automobile market shares.[10] Consider a pair of subcompact cars and an expensive luxury car. If we lower the price of one of the subcompact cars by 10%, then (7.18) says that the percentage change in the demand for the other subcompact car will be the same as the percentage change in the demand for the luxury car (if the other subcompact car demand drops by 20%, then the luxury car demand will also drop by 20%). Such behavior is not very realistic. This IIA behavior stems fundamentally from the i.i.d. assumption on the random-noise terms ξ's of the MNL model. (See Berry [53] for a discussion, and a possible way around these restrictions on cross-price elasticities.)

7.3.4 Stochastic-Demand Functions

A deterministic demand function $d(p)$ can be used to define a stochastic model of demand in a variety of ways. In the stochastic case, we let

[10] If the population is homogeneous, the choice probabilities represent market share, and the MNL can be used to estimate market shares.

$D(p, \xi_t)$ denote the random demand as a function of the price p and a random-noise term ξ_t. The three most common random-demand models are discussed below.

7.3.4.1 Additive Uncertainty

In the additive model, the demand is a continuous random variable of the form

$$D(p, \xi) = d(p) + \xi,$$

where ξ is a zero-mean random variable that does not depend on the price. In this case, the mean demand is $d(p)$, and the noise term ξ shifts the demand randomly about this mean.

Note that this additive disturbance has the property that the elasticity of demand depends on ξ. This follows since

$$\epsilon(p, \xi) = \frac{p}{D(p, \xi)} \frac{\partial D(p, \xi)}{\partial p} = \frac{\epsilon(p)}{1 + \xi / d(p)},$$

where $\epsilon(p) = \frac{p}{d(p)} \frac{\partial d(p)}{\partial p}$ is the deterministic elasticity. So if a realization of ξ is less than zero, the elasticity of demand in the stochastic model is greater than the deterministic elasticity, and if the realization of ξ is greater than zero, it is smaller.

One potential problem with the additive uncertainty model is that demand could be negative if $d(p)$ is small and the variance of ξ is large. For this reason, the additive model should be used with caution in applications where the coefficients of variation for the demand uncertainty is high.

7.3.4.2 Multiplicative Uncertainty

In the multiplicative model, the demand is again a continuous random variable but of the form

$$D(p, \xi) = \xi d(p),$$

where ξ is a nonnegative random variable with mean one that does not depend on the price p. In this case, the mean demand is again $d(p)$, and the noise term ξ simply scales the mean demand by a random factor. For the multiplicative model, the elasticity of demand for any given realization of ξ is the same as the deterministic elasticity, since

$$\epsilon(p, \xi) = \frac{p}{d(p, \xi)} \frac{\partial d(p, \xi)}{\partial p} = \epsilon(p),$$

where again $\epsilon(p)$ is the deterministic elasticity. Thus, the random-noise term does not affect the elasticity of demand; it affects only the magnitude of demand.

Note also that one can also combine the multiplicative and additive uncertainty models, leading to a demand function of the form

$$D(p, \xi) = \xi_1 + \xi_2 d(p),$$

where ξ_1 is a zero-mean random variable and ξ_2 is a nonnegative, unit-mean random variable.

7.3.4.3 Poisson and Bernoulli Uncertainty

Poisson and Bernoulli models of uncertainty are used in the dynamic models of demand discussed in Chapter 5. In the Bernoulli model, $d(p)$ is simply a probability of an arrival in a given period. So $d(p)$ is the probability that demand is one in a period, and $1-d(p)$ is the probability demand is zero. As a result, the mean demand in a period is again $d(p)$, and we can represent the demand as a random function

$$D(p, \xi) = \begin{cases} 1 & \xi \leq d(p) \\ 0 & \xi > d(p) \end{cases},$$

where ξ is a uniform [0,1] random variable.

For example, consider a situation in which the buyer in the period has a reservation price v that is a random variable with distribution $F(\cdot)$. If the firm offers a price of p, they will sell a unit if $v \geq p$, which occurs with probability $1 - F(p)$. This corresponds to setting $d(p) = 1 - F(p)$ above.

In the Poisson model, time is continuous, and $d(p)$ is treated as a stochastic intensity or rate. That is, the probability that we get a unit of demand in an interval of time from t to $t + \delta$ is $\delta d(p) + o(\delta)$ and the probability that we see no demand is $1 - \delta d(p) + o(\delta)$ (all other events have probability $o(\delta)$).

The Poisson and Bernoulli models are useful for several reasons. First, they translate a deterministic demand function directly into a stochastic model, without the need to estimate additional parameters (such as variance). They also are discrete-demand models—as opposed to the continuous demand of the additive and multiplicative models—and more closely match the discreteness of demand in many RM applications. At the same time, the Poisson and Bernoulli models assume a specific coefficient of variation, which may or may not match the observed variability. The additive and multiplicative models, in contrast, allow for different levels of variability in the model, as the complete distribution of the noise term can be specified.

7.3.4.4 Stochastic Regularity

As in the deterministic case, it is useful analytically to make some regularity assumptions about the stochastic demand functions. In particular:

ASSUMPTION 7.6 (STOCHASTIC-DEMAND-FUNCTION REGULARITY) *The variance of demand is uniformly bounded,* $E[|D(p, \xi_t)|^2] \leq K < +\infty$ *for* $p \geq 0$.

This condition is not very restrictive and is required only to ensure that stochastic optimization problems are well behaved.

7.3.5 Rationing Rules

A final demand-modeling issue concerns how capacity is allocated to customers in cases where demand exceeds supply. For example, suppose capacity is 100 units and the firm commits to a fixed price of $10 per unit before knowing the demand realization. If the demand at this price turns out to be 120, then what assumptions do we make about which customers get the capacity and which do not? Do we assume that the capacity is allocated to customers with the highest valuations (thereby increasing the customer surplus), or should we assume that it is allocated randomly—for example, on a first-come, first-serve basis? The rules used for allocating capacity to customers when demand exceeds capacity are called *rationing rules* in economics.

There are two classical rationing rules: (1) The *efficient-rationing rule* (also called *parallel rationing*), in which it is assumed that units are allocated to customers with the highest valuations, and (2) the *proportional-rationing rule* (also called *randomized rationing*), in which it is assumed that capacity is allocated randomly, so the allocation is independent of the customers' valuations. While the former is more efficient from a consumer surplus standpoint, it is difficult to achieve in most posted-price settings (though some types of auctions implement it very well).

In quantity-based RM applications the most natural assumption is the proportional-rationing rule because when a given product is open, it is normally purchased on a first-come first-served basis. Therefore, provided there is no correlation between valuations and order of arrival, the inventory is sold independent of valuations.

7.4 Notes and Sources

Kreps [313] provides a comprehensive and readable treatment of the classical rational theory of consumer choice, including preference relations, utility theory, and choice under uncertainty. See also the microeconomics text of Mas-Collel et al. [365].

Random-utility models originated with the early work of the mathematical psychologist Thurston [511, 510] and were later formalized by economists, most notably Manski [358] and McFadden [372, 373]. (See also the edited volume by Manski and McFadden [357].) The limitations of the MNL as a model for transportation demand are discussed in detail by Oum [412]. The Gumbel distribution, which plays a central role in the MNL, is one of the distributions of extremes examined in Gumbel [229].

Kamakura and Russell [286], Chintagunta [117], and Allenby, Arora and Ginter [8] are some marketing-science papers that use the finite-mixture logit models. The finite-mixture and random-coefficient models are said to be heterogeneous in preferences; that is, customers use the *same* choice model but have different preferences (for example, use different coefficients) within that choice model. Another source of heterogeneity—called *structural heterogeneity*—is when customers in different segments use fundamentally different decision processes in making their purchase decisions. Such structural heterogeneity is studied in Kannan and Wright [287] and Kamakura, Kim, and Lee [285]. Finally, Dirichlet distributions have been used to model heterogeneity in brand-choice behavior (Fader and Lattin [179]; Jain, Bass, and Chen [266]).

An excellent comprehensive text on both the theory and application of discrete-choice models for demand estimation is Ben-Akiva and Lerman [48]. See also the book by Anderson et al. [16] for another good text on discrete-choice theory and economic-modeling applications of the theory.