Models of Demand *

Garrett J. van Ryzin

November 1, 2010

Contents

*Portion adapted from Talluri and van Ryzin [37].

1 Introduction

Demand models are essential for analytical pricing. They predict how a firm's pricing actions will influence demand for its products and services, which in turn determines revenues and profits. Here we review the most important demand models used in pricing practice and survey approaches to estimating these models.

The most widely used models of demand assume customers are *rational* decision makers who intelligently alter when, what, and how much to purchase to achieve the best possible outcome for themselves. This is a quite plausible assumption. Moreover, an important consequence of this rationality assumption is that customer behavior can be "predicted" by treating each customer as an agent that optimizes over possible choices and outcomes. Optimization theory can then be used to model their behavior. These are the most widespread models used in pricing practice and the focus of this chapter. Other theories of customers from behavioral science depart from this classical rational view and examine important biases in buying decisions. These theories are covered in Chapter [OZER AND ZHENG] of this handbook and are also discussed briefly below.

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, etc. – we begin by looking at models of individual-choice behavior. When added up, these individual purchase decisions determine aggregate demand, so we next examine aggregate-demand functions and their properties. Finally, we survey how such models can be estimated from data.

2 Consumer Theory

For completeness, here we briefly review the fundamentals of classical consumer theory from economics, highlighting the concepts and assumptions that are important for operational demand modeling. A complete treatment of consumer theory is given in Chapter [WEBER] of this volume.

2.1 Choice and Preference Relations

What do we mean by choice? Formally, given two alternatives, a *choice* corresponds to an expression of preference for one alternative over another. Here, "alternatives" may refer to different products, different quantities of the same product, bundles of different products or various uncertain outcomes (such as buying a house at the asking price versus waiting and bidding in an auction against other buyers). Similarly, given n alternatives, choice can be defined in terms of the preferences expressed for all pairwise comparisons among the n alternatives.

The mathematical construct that formalizes this notion is a *preference relation*. Customers are assumed to have a set of *binary preferences* over alternatives in a set X; given any two alternatives x and y in X, customers can rank them and clearly say they prefer one over the other or are indifferent between them. This is represented by the notation $x \succeq y$. A customer strictly prefers x to y, denoted $x \succ y$, if he prefers x to y, but does not prefer y to x (that is, he is not indifferent between the two alternatives). Consider a complete set of all such pairwise binary preferences between alternatives in X with the following two properties:

Asymmetry If x is strictly preferred to y, then y is not strictly preferred to x.

Negative transitivity If x is not strictly preferred to y and y is not strictly preferred to z, then x is not strictly preferred to z.

Such a binary relation is called a preference relation. While asymmetry is a very plausible assumption, negative transitivity is not completely innocuous, and it is possible to construct natural examples of preferences that violate it. Still, preference relations form the classical basis for modeling customer choice.

The following is a common example of a preference relation:

Example 1 *(*Address model*) Suppose we have n alternatives and each alternative has m attributes that take on real values. Alternatives can then be represented as n points, $z_1, \ldots, z_n$, in $\Re^m$, which is called the* attribute space*. For example, attributes could include color, size, indicators for features and price.*

Each customer has an ideal point ("address") $y \in \Re^m$, reflecting his most preferred combination of attributes (such as an ideal color, size and price). A customer is then assumed to prefer the product closest to his ideal point in attribute space, where distance is defined by a metric ρ on $\Re^m \times \Re^m$ (such as Euclidean distance). These distances define a preference relation, in which $z_i \succ z_j$ if and only if $\rho(z_i, y) < \rho(z_j, y)$; that is, if z_i is "closer" to the ideal point y of the customer.

2.2 Utility Functions

Preference relations are intimately related to the existence of utility functions. Indeed, we have the following theorem (See Kreps [28] for a proof.):

Theorem 1 *If X is a finite set, a binary relation $\succ$ is a preference relation if and only if there exists a function $u : X \to \Re$ (called a **utility function**), such that*

$$x \succ y \quad \text{iff} \quad u(x) > u(y).$$

Intuitively, this theorem follows because if a customer has a preference relation, then all products can be ranked (totally ordered) by his preferences; a utility function then simply assigns a value corresponding to this ranking. Intuitively, one can think of utility as a cardinal measure of "value," though in a strict sense its numerical value need not correspond to any tangible measure. Theorem 1 applies to continuous sets X (such as travel times or continuous amounts of money) as well under mild regularity conditions, in which case the utility function $u(\cdot)$ is then continuous. Continuing our previous example:

Example 2 *Consider the address model of Example 1. Theorem 1 guarantees an equivalent utility maximization model of choice. In this case, it is easy to see that for customer y the continuous utilities*

$$u(z) = c - \rho(z, y),$$

where c is an arbitrary constant, produce the same preferences as the address model.

2.3 Consumer Budgets, Demand and Reservation Prices

Demand can be derived once customers' preferences are known. Given a customer's preferences for bundles of n goods, a vector of market prices $\mathbf{p} = (p_1, \ldots, p_n)$ for these goods, and a level of monetary wealth (or *income*) w, we simply ask how our customer would "spend" his wealth? Let x_i denote the quantity of each good i consumed. We assume x_i are continuous quantities and our customer has a continuous utility function $u(\mathbf{x})$. Let $\mathbf{x} = (x_1, \ldots, x_n)$. The *consumer budget problem* can then be formulated as

$$\begin{aligned} \max \quad & u(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{p}^\top \mathbf{x} \leq w \\ & \mathbf{x} \geq 0. \end{aligned} \tag{1}$$

In other words, a customer's purchase quantities maximize their total utility subject to their limited wealth (income). The resulting optimal $\mathbf{x}^*$ is a customer's demand vector for the n goods at price vector $\mathbf{p}$. [1]

2.4 Reservation Prices

A *reservation price*, denoted v_i, is the monetary amount a customer is willing to give up to acquire an additional marginal amount of good i. Reservation prices are also referred to as the customer's *willingness to pay.* One can show from the first-order conditions of the budget problem that if $x_i^* > 0$, then $v_i = p_i$. Thus, a customer's reservation price for goods that are *currently consumed* is simply the market price. The reasoning is intuitive; if our customer valued consuming an additional marginal amount of good i at strictly more than its market price, then he would be able to increase his utility by marginally reducing consumption of other goods and increasing his consumption of good i. Since our customer is maximizing utility, this cannot occur.

On the other hand, for goods i that are not being consumed, $x_i^* = 0$ and it follows from the first-order conditions of the budget problem that $v_i \leq p_i$. In other words, the customer's reservation price for initial consumption of good i is less than its current market price. Moreover, the customer would only change his allocation and buy good i if its price p_i dropped below his reservation price v_i.

As a practical matter, this formal theory of reservation prices is less important than the informal concept—namely, that a reservation price is the maximum amount a customer is willing to pay for an additional unit of good i and to entice a customer to buy good i, the price must drop below his reservation price. Still, the analysis highlights the important fact that reservation prices are not "absolute" quantities. They depend on customers' preferences, wealth, their current consumption levels, and the prices of other goods the customers may buy; change one of these factors, and customers' reservation price may change.

2.5 Preference for Stochastic Outcomes

Many choices involve uncertain outcomes, such as buying insurance, making investments or even eating at a new restaurant. How do customers respond

[1] The optimal consumption vector x^* may not be unique, in which case $x * (\mathbf{p})$ is a demand correspondence rather than a demand function; see Chapter [WEBER] in this volume for more details.

to such uncertainties? The theory of choice under uncertainty is a deep and extensive topic. Here, we outline only the main concepts.

Consider again a discrete set of n alternatives, $X = \{x_1, \ldots, x_n\}$. Let $\mathcal{P}$ be the class of all probability distributions $P(\cdot)$ defined on X. That is, $P \in \mathcal{P}$ is a function satisfying $\sum_i P(x_i) = 1$ and $P(x_i) \geq 0$ for $i = 1, \ldots, n$. One can think of each P as a "lottery," the outcome of which determine which one of the alternatives the customer gets according to the distribution P.

What can we say about a customer's preference for lotteries? Specifically, when can we say that for any two lotteries P_1 and P_2, customers "prefer" one over the other (denoted by $P_1 \succ P_2$)?

To answer this question we need to make some assumptions on customer preferences. First, we will assume there exists a preference relation $\succ$ on the n different outcomes x_i as before. Second, for any two lotteries P_1 and P_2, consider a compound lottery parameterized by α as follows: i) A coin is flipped with probability of heads equal to α; ii) If the coin comes up heads, the customer enters lottery P_1, otherwise the customer enters lottery P_2. Denote this compound lottery by $\alpha P_1 + (1-\alpha)P_2$. Note this compound lottery is also contained in the set $\mathcal{P}$ (i.e., $\mathcal{P}$ is a convex set). We then require the following consistency properties on a customer's preference for lotteries:

Substitution axiom For all P_1, P_2, and P_3 in $\mathcal{P}$ and all $\alpha \in (0, 1]$, if $P_1 \succ P_2$, then $\alpha P_1 + (1-\alpha)P_3 \succ \alpha P_2 + (1-\alpha)P_3$.

Continuity axiom For all P_1, P_2, and P_3 in $\mathcal{P}$ with $P_1 \succ P_2 \succ P_3$, there exist values $\alpha \in (0, 1)$ and $\beta \in (0, 1)$ such that $\alpha P_1 + (1-\alpha)P_3 \succ P_2 \succ \beta P_1 + (1-\beta)P_3$.

Roughly, the first axiom says that substituting a gamble that produces a strictly preferred collection of outcomes in the compound lottery should be preferred. The second axiom says that if a customer strictly prefers one gamble to another, then he should be willing to accept a sufficiently small risk of an even worse outcome to take the preferred gamble. We then have:

Theorem 2 *A preference relation on the lotteries $\mathcal{P}$ exists that satisfies the substitution and continuity axioms if and only if there exists a utility function $u(\cdot)$ such that $P_1 \succ P_2$ if and only if*

$$\sum_{i=1}^{n} u(x_i)P_1(x_i) > \sum_{i=1}^{n} u(x_i)P_2(x_i).$$

That is, if and only if the expected utility from lottery P_1 exceeds the expected utility of lottery P_2. In addition, any two utility functions u and u' satisfying *the above must be affine transformations of each other; that is,*

$$u(x) = cu'(x) + d,$$

for some real $c > 0$ and d.

This result is due to von Neumann and Morgenstern [41] and the function $u(x)$ above is known as the *von Neumann-Morgenstern utility*. Essentially, this result allows us to extend utility as a model of customer preference to the case of uncertain outcomes, with expected utility replacing deterministic utility as the criterion for customer decision making.

2.6 Risk Preferences

An important special case of expected-utility theory is when outcomes correspond to different levels of wealth and lotteries correspond to different gambles on a customer's ending wealth level. We assume the wealth levels are continuous and that the customer has preferences for wealth that satisfy the conditions of Theorem 2. Also, assume the lotteries are now continuous distributions F on $\Re$.[2]

Consider now any given lottery F (a distribution on possible wealth outcomes) and μ_F denote the mean of the distribution. A customer is said to have *risk-averse preferences* if he prefers the certain wealth μ_F to the lottery F itself for all possible lotteries F. That is, the customer always prefers the certainty of receiving the expected wealth rather than a gamble with the same mean. The customer is said to have *risk-seeking preferences* if he prefers the gamble F to the certain outcome μ_F for all F. Finally, he has *risk-neutral preferences* if he is indifferent between the lottery F and the certain reward μ_F.[3] Risk preferences are linked directly to concavity or convexity of the customer's utility function as shown by the following theorem:

Theorem 3 *A customer's preference $\succ$ for lotteries exhibits risk-aversion (risk-seeking) behavior if and only if his von Neumann-Morgenstern utility*

[2]The extension of Theorem 2 to the continuous case requires some additional technical conditions that are beyond the scope of this chapter. See Kreps [28].

[3]Note that a customer's preferences may not fall into any of these three categories. For example, many customers take out fire insurance, preferring a certain loss in premium payments every year to the gamble between making no payments but potentially loosing their house, yet simultaneously play their local state lottery, which has an expected loss but provides a small probability of a large wealth pay-off. Such behavior violates strict risk preference.

function $u(w)$ is concave (convex). His preference is risk-neutral if and only if $u(w)$ is affine.

The reason is quite intuitive; with a concave utility function for wealth, a customer gains less utility from a given increase in wealth than he loses in utility from the same decrease in wealth. Hence, the upside gains produced by the volatility in outcomes do not offset the downside losses, and customers therefore prefer the certain average to the uncertain outcomes of the lottery. Since most customers have a decreasing marginal utility for wealth, risk aversion is a good assumption in modeling customer behavior.

Still, the concept of risk aversion has to be addressed with care in operational modeling. While it is true that most customers are risk-averse when it comes to *large* swings in their wealth, often the gambles we face as customers have a relatively small range of possible outcomes relative to our wealth. For example, a customer may face a price risk in buying a CD or book online. However, the differences in prices for such items are extremely small compared to total wealth. In such cases, the utility function may be almost linear in the range of outcomes affecting the decision, in which case the customer exhibits approximately risk-neutral preference.[4] Similar statements apply to firms. Generally, they are risk-averse too, but for decisions and gambles that involve "small" outcomes relative to their total wealth and income, they tend to be approximately risk-neutral. Hence, risk-neutrality is a reasonable assumption in operational models and, indeed, is the standard assumption in pricing practice.

3 Disaggregate Demand Models

With the basic concepts of consumer theory in hand, we next look at approaches for modeling individual customer purchase decisions. Such models may be used on their own to predict individual purchase instances (e.g., the likelihood a customer buys when visiting a web site) or as building blocks in aggregate demand models, which are discussed in the subsequent section. With the increased availability of individual customer level data and the trend toward making pricing more targeted, such disaggregate models of demand are increasingly being used in pricing practice.

[4]Formally, one can see this by taking a Taylor series approximation of the utility function about the customer's current wealth w; the first-order approximation is affine, corresponding to risk-neutrality.

3.1 An Overview of Random-Utility Models

The disaggregate demand models most commonly used are based on a probabilistic model of individual customer utility. This framework is useful for several reasons. First, probabilistic models can be used to represent heterogeneity of preference among a population of customers. They can also model uncertainty in choice outcomes due to the inability of the firm to observe all the relevant variables (other alternatives, their prices, the customer's wealth, and so on) or situations where customers exhibit *variety-seeking behavior* and deliberately alter their choices over time (movie or meal choice, for example). Finally, probabilistic utility can model customers whose behavior is inherently unpredictable—that is, customers who behave in a way that is inconsistent with well-defined preferences and who at best, exhibit only a probabilistic tendency to prefer one alternative to another. (See Luce [30].)

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 *random* component, ξ_j (often assumed to have zero or constant mean for all j). Therefore,

$$U_j = u_j + \xi_j, \tag{2}$$

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

$$P_j(S) = P(U_j \geq \max\{U_i : i \in S\}). \tag{3}$$

In other words, the probability of selecting alternative j is the probability that it has the highest utility among all the alternatives in the set S. The set S (called the *choice set*) is a collectively exhaustive, mutually exclusive set of all the feasible alternatives considered by a customer when making a choice decision. It varies based on which alternatives a firm makes available (the *offer set*), the set of outside alternatives (competitors' offer sets or substitute options) and which subset of these internal and external alternatives a customer actually considers when making a choice (this may vary by customer segment). Identifying appropriate choice sets is a challenging modeling task. In operational modeling, a common simplification is to aggregate all competitor and substitute alternatives into a single generic "outside" alternative or "no purchase" alternative.

The representative component u_j is often modeled as a function of the relevant attributes of alternative j. A common assumption is the *linear-in-parameters* model

$$u_j = \boldsymbol{\beta}^\top \mathbf{x}_j, \tag{4}$$

where β is a vector of weights and $\mathbf{x}_j$ is a vector of attribute parameters for alternative j, which could include factors such as price, measures of quality and indicator variables for product features. [5] Variables describing characteristics of the customer (segment variables) can also be included in $\mathbf{x}_j$.[6]

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 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. However, as a practical matter, certain assumptions on the random utilities lead to simpler models than others. We look at a few of these special cases next.

3.2 Binary Probit

If there are only two alternatives to choose from (e.g., buying or not buying a product), the error terms $\xi_j, j = 1, 2$, are normally distributed random variables with mean zero and the difference in the error terms, $\xi = \xi_1 - \xi_2$, has variance σ^2, then the probability that alternative 1 is chosen is given by

$$P(\xi_2 - \xi_1 \leq u_1 - u_2) = \Phi(\frac{u_1 - u_2}{\sigma}), \tag{5}$$

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 (i.e., it can be viewed as resulting from the sum of a large number of random disturbances), the resulting probabilities do not have a simple closed-form solution.

[5] Attribute parameters can also include transforms of attribute values, such as the natural log or square root.

[6] 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.

3.3 Binary Logit

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

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

where $\mu > 0$ is a scale parameter. Here ξ has a mean zero and variance $\frac{\pi^2}{3\mu^2}$. 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

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

3.4 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

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

where μ is a positive scale parameter and η is the mode of the distribution. The mean and variance of ξ_j are

$$E[\xi_j] = \eta + \frac{\gamma}{\mu}, \qquad \mathrm{V}ar[\xi_j] = \frac{\pi^2}{6\mu^2},$$

where γ is Euler's constant ($= 0.5772\ldots$). The *standard Gumbel distribution* has $\eta = 0$ and $\mu = 1$, and since utility can be shifted and scaled without altering choice outcomes one often assumes disturbances have a standard Gumbel distribution in the MNL.

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 is also a Gumbel random variable. Also, if two random variables ξ_1 and ξ_2 are independent and Gumbel distributed, then $\xi = \xi_1 - \xi_2$ has a logistic distribution. (See Ben-Akiva and Lerman [1] and Train [38] for details.)

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

$$P_j(S) = \frac{e^{\mu u_j}}{\sum_{i \in S} e^{\mu u_i}}. \tag{6}$$

If $\{u_j : j \in S\}$ has a unique maximum and $\mu \to \infty$, 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 $\mu \to 0$, 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 among the alternatives in S. Hence, the MNL can model behavior ranging from deterministic utility maximization to purely random choice.

Though widely used as a model of customer choice, the MNL 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

$$\frac{P_i(S)}{P_j(S)} = \frac{P_i(T)}{P_j(T)}. \tag{7}$$

Equation (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 then adding a new alternative reduces the probability of choosing similar alternatives more than dissimilar alternatives, violating the IIA property. A famous example illustrating this point is the "blue-bus/red-bus paradox," (Debreu [10]):

Example 3 *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*

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

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

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

However, as bus's 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

$$\begin{aligned} P_{car}(T) &= \frac{1}{2} \\ P_{blue\ bus}(T) = P_{red\ bus}(T) &= \frac{1}{4}. \end{aligned}$$

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."

Despite this deficiency, the MNL model is widely used in marketing. (See Guadagni and Little [15].) It has also seen considerable application in estimating travel demand. (See Ben-Akiva and Lerman [1].) 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.

3.5 Finite-Mixture Logit Models

In the basic MNL model with linear-in-attribute utilities, the coefficients $\boldsymbol{\beta}$ in (4) 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 have 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 we cannot identify which customers belong to each segment.

Assume that customers within each segment follow a MNL model with identical parameters and that customers have a certain probability of belonging to a finite number of segments (called *latent segments*), 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. 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

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

The mixed logit is an important generalization of the MNL. Indeed, McFadden and Train [32] show that any discrete choice model derived from random utility can be approximated arbitrarily closely by an appropriate finite mixing distribution. One can estimate the coefficients of the model ($\boldsymbol{\beta}$ and $q_l,\ l = 1, \ldots, L$) using, for example, maximum-likelihood methods as discussed below. This model often provides better estimates of choice behavior than the standard MNL model, at the expense of a more complicated estimation procedure. Also, the more segments included in the model, the more parameters there are to estimate and more data is required to get high-quality estimates.

3.6 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 among 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

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

But $\boldsymbol{\beta}$ 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 $\mathbf{b}$ 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 $\zeta_m,\ m = 1,\ldots,M$ 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

$$\nu_j \quad = \quad [\sum_{m=1}^{M} x_{mj}\sigma_m\zeta_m] + \xi_j, \quad j = 1,\ldots,n. \tag{8}$$

Then a customer's random utility is given by

$$U_j(\boldsymbol{\nu}) = \mathbf{b}^\top \mathbf{x}_j + \nu_j, \quad j = 1,\ldots,n, \tag{9}$$

where ν_j is given by (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 importantly, they are no longer Gumbel distributed.

4 Aggregate Demand Models

It is often easier to model and estimate aggregate demand rather than individual customer-choice decisions. Depending on the model, this 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 the specification. In this section, we look at some commonly used aggregate-demand models.

Before doing so, however, note that aggregate demand models can be formed from disaggregate demand models simply by "adding up" the effect of many individual consumer choice decisions. For example, suppose we know that the probability that a given consumer purchases an alternative j from a set of alternatives S is $P_j(S)$. This probability could come from a MNL model, a mixed logit model or a more complex discrete choice model. Suppose further that we know that N such customers will face the same choice. Then the aggregate expected demand for j is simply $D_j(S) = NP_j(S)$. Likewise, as we show below, aggregate demand models can often be interpreted as stemming from an underlying disaggregate demand model that is multiplied by a "market size" to produce aggregate demand. So in many ways, the distinction between aggregate and disaggregate models is not sharp - both are ultimately trying to explain the same real-world demand. The choice of which approach to use depends on the data available, the pricing decision being addressed and, to a very real extent, the preference of the model builder.

4.1 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) must be constrained further to ensure they produce non-negative demand.

4.1.1 Regularity

It is often convenient to assume the following regularity conditions about the demand function:

Assumption 1 (Regularity: Scalar Case)
(i) The demand function is continuously differentiable 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 technical complications in both analysis and numerical optimization. For example, consider a linear demand model (defined formally in Section 4.3.1)

$$d(p) = a - bp, \quad p \in \Omega_p = [0, a/b]. \tag{10}$$

This is trivially differentiable on Ω_p, is strictly decreasing if $b > 0$, is non-negative 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 $p^0 = \frac{a}{2b}$.

4.1.2 Reservation-Prices and Demand Functions

Demand functions can naturally be interpreted in terms of a reservation price model in which 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 does not purchase. Moreover, customers buy only one unit of the product. A customer's reservation price is specific to each individual and is normally private information unknown to the firm, though one may estimate a distribution of reservation prices across a population of customers. Let the probability that a customer's reservation price is below p be given by $F(p) = P(v \leq p)$. The derivative of $F(p)$ is denoted $f(p) = \frac{\partial}{\partial p}F(p)$.

Using this model, it is often convenient to express the demand function in the form

$$d(p) = N(1 - F(p)), \tag{11}$$

where $F(p)$ is the reservation price distribution and N is interpreted as the *market size*. $1-F(p)$ is then the fraction of customers willing to buy at price p. For example, consider again the linear-demand function (10). This can be written in the form (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.

4.1.3 Elasticity of Demand

The *price elasticity* of demand is the relative change in demand produced by a unit 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 (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 1 shows a sample of estimated elasticities for common consumer products. This table distinguishes between short and long run elasticities for certain products. For example, in the short-term consumers can not do much about an increase in natural gas prices; they might slightly decrease consumption by lowering thermostats, using supplementary heating sources, etc. Such behavior will not significantly affect demand for natural gas. However, a long-term increase in natural gas prices could easily cause consumers to replace gas furnaces with oil heating or other alternative energy sources, causing a much more significant loss in demand. This is consistent with the numbers: short term elasticity for gas is much lower than the long-term elasticity (0.1 vs. 0.5). While many factors affect elasticity, these estimates give some sense of the relative magnitudes of elasticities.

4.1.4 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 in the range of the demand function (domain of the inverse) implies a unique choice of price $p(d)$. This is often a useful transformation, as the resulting revenue and profit functions are usually simpler to work with, both analytically and computationally.

Equation (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 (10) is

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

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

4.1.5 Revenue Function

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

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

Table 1: Estimated elasticities (absolute values) for common products.[a]

Product	$\lvert\epsilon(p)\rvert$
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 [16], collected from various econometric studies.

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 (10) has a revenue function

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

For most pricing models, it is desirable for this revenue function to be concave or at least unimodal, as in the linear example above. This leads to optimization problems that are numerically well behaved.

4.1.6 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

$$\begin{aligned} J(d) &\equiv \frac{\partial}{\partial d} r(d) \\ &= p(d) + dp'(d). \end{aligned} \tag{12}$$

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[7]

$$J(p) \equiv J(d(p)) = p + d(p)\frac{1}{d'(p)}. \tag{13}$$

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.[8]

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

$$J(p) = p - \frac{1}{\rho(p)}, \tag{14}$$

where $\rho(p) \equiv f(p)/(1 - F(p))$ is the *hazard rate* of the distribution $F(p)$. To illustrate, consider the marginal revenue of the linear-demand function of (10) as a function of d,

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

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

[8] 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.)

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

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

4.1.7 Revenue Maximization

Maximizing revenue and profit is a central problem in pricing analysis. Assuming the maximizer is an interior point of the domain Ω_p, 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 (13) that since $J(p) = p(1 + \frac{d}{p}\frac{\partial p}{\partial d})$, $\frac{\partial d}{\partial p} < 0$ (from Assumption 1, part (ii)), and $\frac{\partial d}{\partial p}\frac{p}{d} = \epsilon(p)$ is the price elasticity, we have

$$J(p) = p\left(1 - \frac{1}{|\epsilon(p)|}\right). \tag{15}$$

Thus, marginal revenue is positive (revenues are increasing in demand, decreasing in price) if demand is elastic at p (that is, if $|\epsilon(p)| > 1$), and marginal revenue is negative (revenues are decreasing in demand, increasing in price) 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.

When are these first-order conditions sufficient? If the revenue function is unimodal,[9] this suffices to ensure that the first order conditions determine an optimal price. The following sufficient conditions on the reservation-price distribution ensure unimodality (or strict unimodality) of the revenue function (see Ziya et al. [45]):

[9] 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 increasing on $[a, x^*]$ and $f(x)$ is decreasing on $[x^*, b]$; it is *unimodal* 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 $[x^*, b]$

Proposition 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) $r(d) = dp(d)$ is (strictly) concave on $d \in \Omega_d$.
(ii) $r(p) = pd(p)$ is (strictly) concave on $p \in \Omega_p$.
(iii) The absolute value of the elasticity $|\epsilon(p)| = p\rho(p)$ is (strictly) increasing on $p \in \Omega_p$ and $\sup_{p \in \Omega_p} |\epsilon(p)| > 1$.

Then the revenue function $r(p) = pd(p) = pN(1 - F(p))$ is (strictly) unimodal on Ω_p (equivalently, the revenue function $r(d) = p(d)d$ is (strictly) unimodal on $\Omega_d = [N(1 - F(p_1)), N(1 - F(p_2))]$. Moreover, the above three conditions are (respectively) equivalent to:
(i) $2\rho(p) \geq -\frac{f'(p)}{f(p)}$ for all $p \in \Omega_p$.
(ii) $\frac{2}{p} \geq -\frac{f'(p)}{f(p)}$ for all $p \in \Omega_p$.
(iii) $\rho(p) + \frac{1}{p} \geq -\frac{f'(p)}{f(p)}$ for all $p \in \Omega_p$
with strict inequality replacing equality for strict unimodality.

These conditions are intuitive. The first two state that if the revenue function is concave when parameterized either in terms of quantity (demand) or price, this implies unimodality of $r(p)$. For the third condition – that the magnitude of the elasticity, $|\epsilon(p)|$ is increasing – recall the revenue is increasing in price when $|\epsilon(p)| < 1$ and decreasing in price when $|\epsilon(p)| > 1$; hence, it is clear that there is a revenue maximizing price if elasticity is monotone increasing. Ziya et al. [45] show there are demand functions that satisfy one condition but not the others, so the three conditions are indeed distinct.

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.[10]

[10]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.

4.1.8 Variable Costs and Profit Maximization

The prior results on revenue maximization extend to the case where there are variable costs and the firm seeks to maximize profits (revenues minus costs). Let the cost of supplying d units of demand be denoted $c(d)$. We will assume this cost function is convex and continuously differentiable. Let $c'(d)$ denote the marginal cost (derivative of $c(\cdot)$). In this case, the optimization problem is

$$\max_{d}\{r(d) - c(d)\}, \tag{16}$$

and the first-order conditions for the optimal demand level are

$$J(d) = c'(d). \tag{17}$$

That is, the optimal volume of demand (and associated price) occurs at the point where marginal revenue equals marginal cost. For the constant marginal cost case $c(d) = cd$, using (15) and rearranging, this optimality condition can be written in the convenient form:

$$\frac{p-c}{p} = \frac{1}{|\epsilon(p)|}. \tag{18}$$

The left hand side is simply the relative profit margin (profit as a fraction of price), and the left hand side is the inverse of the elasticity. When costs are zero, the relative profit margin is always one, so the optimal price occurs when the elasticity is one, as we observed above. When marginal cost are positive, however, the relative profit margin will be less than one, so the optimal price occurs at a point where elasticity is greater than one (the elastic region).

4.2 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, $\boldsymbol{\Omega}_p$ 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 2 (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: $0 \leq d_j(\mathbf{p}) < +\infty, \ \forall \mathbf{p} \in \mathbf{\Omega}_p$.
(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 to be defined below in Section 4.3, 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 $\mathbf{p}(\mathbf{d})$. Again, for numerical optimization, it is desirable if this revenue function is jointly concave.

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).

4.3 Common Demand Functions

Table 2 summarizes the most common demand functions and their properties. All these functions satisfy the regularity conditions in Assumptions 1 and 2, except for the constant-elasticity-demand function, which does not satisfy part (v) of either assumption as explained below.

Unfortunately, there are no hard and fast rules about which demand function to use in a given application. Any of the functions in Table 2 will work well in modeling modest price adjustments around a fixed price point; in this case only the local properties (slope) of the function matter. The models show greater difference when applied to a wider range of prices. Some, like the log-linear and logit model, have more reasonable properties when extrapolated to extremely high or low prices. Estimation concerns can also dictate the choice; the log-linear model, for example, is poorly suited to sparse demand applications since zero-demand observations cause problems when using the log-demand transformation typically required to estimate the model via regression. Simplicity and ease of use matter as well; the linear demand model remains quite popular in practice because it is both simple analytically and easy to estimate using regression. The intended use of the model is also important. For example, constant elasticity models are popular for econometric studies, but assuming elasticity is constant for all prices leads to highly unrealistic behavior in optimization models. In short, the best model to use in any given application depends on the characteristics of the demand, its ease of use and the intended purpose of the modeling effort; ultimately, it is more an engineering than a scientific choice.

4.3.1 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. 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

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

Table 2: 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)	e^{a-bp}	$\frac{1}{b}(a-\ln(d))$	$\frac{d}{b}(a-\ln(d))$	$\frac{1}{b}(a-1-\ln(d))$	pb	$\frac{1}{b}$
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	$N\frac{e^{-bp}}{1+e^{-bp}}$	$\frac{1}{b}\ln(\frac{N}{d}-1)$	$\frac{d}{Nb}\ln(\frac{N}{d}-1)$	$\frac{1}{b}\left(\ln(\frac{N}{d}-1)-\frac{N}{d-N}\right)$	$\frac{bp}{1+e^{-bp}}$	No closed-form formula

Definitions: $d(p)$ = demand function, $p(d)$ = inverse demand function, $r(d)$ = revenue function ($r(d) = dp(d)$), $J(d)$ = marginal revenue ($J(d) = \partial r(d)/\partial d$), $\|\epsilon(p)\|$ = elasticity ($\|\epsilon(p)\| = \|\frac{p}{d}\frac{\partial p(d)}{\partial d}\|$), p_0 is the revenue maximizing price.

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 (See Horn and Johnson [22], Vives [40].):

$$b_{ii} < 0 \quad \text{and} \quad |b_{ii}| > \sum_{j \neq i} |b_{ij}|, \quad i = 1, \ldots, n. \tag{19}$$

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

$$b_{jj} < 0 \quad \text{and} \quad |b_{jj}| > \sum_{i \neq j} |b_{ij}|, \quad j = 1, \ldots, n. \tag{20}$$

Equation (20) 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).

4.3.2 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 natural 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 when sales volumes are low.

The multidimensional log-linear form is

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

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

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

so again the log-linear model can be estimated easily from data using linear regression provided zero sales instances are not frequent.

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 (19) or (20) to check that $\mathbf{B}^{-1}$ exists.

4.3.3 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 (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 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 it is less problematic.

The multiproduct constant elasticity model is

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

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

4.3.4 Logit Demand

The logit demand function is based on the MNL model of Section 3.4. Since utility is ordinal, without loss of generality we can assume the no-purchase utility $u_0 = 0$. The choice probabilities are then given by (6) with the no-purchase alternative having a value $e^{u_0} = 1$.

As mentioned, it is common to model the representative component of utility u_j as a linear function of several known attributes including price. Assuming u_j is linear in price and interpreting the choice probabilities as fractions of a population of customers of size N leads 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 $\Omega_p = [0, +\infty)$. The inverse-demand function is

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

Logit demand models have a desirable "S"-shape that many practitioners find appealing and intuitively plausible; the function in fact can be interpreted as resulting from a population of reservation prices that is approximately normally distributed around a mean value (the logistical distribution in the error term of the utility approximates a normal random variable).

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). 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

$$\begin{aligned} \epsilon_{ij}(\mathbf{p}) &= \frac{\partial \ln d_i(p)}{\partial \ln p_j} \\ &= -p_j b_j \frac{e^{-b_j p_j}}{1 + \sum_{k=1}^{n} e^{-b_k p_k}}. \end{aligned} \tag{21}$$

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.[11] 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 (21) says that the percentage change in the demand for the other subcompact car will be the same as the percentage

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

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 [2] for a discussion, and a possible way around these restrictions on cross-price elasticities.)

4.4 Stochastic-Demand Functions

At disaggregate levels of modeling - e.g., modeling sales in a given store on a given day or an individual on-line shopper's decision to buy or not buy - demand outcomes are inherently uncertain. In such cases, stochastic demand models are required. We can convert a deterministic demand function $d(p)$ into a stochastic model of demand in a variety of ways. In the stochastic case, we let $D(p, \xi_t)$ 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.

4.4.1 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.

4.4.2 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.

4.4.3 Poisson and Bernoulli Uncertainty

In the Bernoulli model, $d(p)$ is the 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 $[t, t+\delta)$ 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 pricing 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.

4.5 Subrational Behavior Models

While rational behavior is the standard assumption underlying most of the theory and practice of RM, it is far from being completely accepted as a model of how an actual customer behaves. Indeed, much of the recent work in economics and customer behavior has centered on explaining observed, systematic deviations from rationality on the part of customers. These theories are covered in detail in Chapter [OZER AND ZHENG] of this handbook. Here we only give a brief overview of the main ideas and their implications for demand modeling.

Kahneman and Tversky [26, 25] famously showed that people exhibit consistent biases when faced with simple choices in an experimental setting. Their key insight is that most individuals tend to evaluate choice in terms of losses and gains from their status quo wealth, rather than evaluating choices in terms of their terminal wealth as in classical utility theory. They show a tendency toward "loss aversion" rather than risk aversion, and they have a strong preference for certainty of outcomes when evaluating choices. Other experiments revealed that people put a much higher value on a product they already own than one that they don't own because giving up a product they have feels like a loss. This is known as the *endowment effect*. Another bias people exhibit is due to what is called *mental accounting*, in which customers tend to evaluate gains and losses for different categories of goods differently because they have "mental budgets" for each category of goods.

Subrational behavioral theory has potentially profound implications for tactical demand modeling and pricing optimization. For example, if customers base their purchase decisions on perceived losses and gains, imagine how this might affect retail pricing decisions. For instance, suppose sales of a product at its initial price are low. Currently, most tactical pricing model would interpret this as a clear pricing error; we simply priced too high initially. But under prospect theory, such a high price might actually help because once the price is reduced, customers will then perceive the reduction as a gain. Indeed, it may be optimal to set the initial price very high - or leave it high longer - because then the perceived gain once the price is lowered is that much greater. Such notions are of course part and parcel of the general marketing toolkit, but today they are applied quite crudely. The potential is to deploy these same concepts with the sort of precision and detail made possible by a model-based pricing system.

Do such findings mean that current demand models based on rational behavior are obsolete? Not exactly. In a gross sense, people do tend to behave in accordance with rationality assumptions; they buy more when prices are lower and less when prices are high. However, what this behavioral theory and supporting laboratory and real-world observation shows is that the axioms of rational behavior, plausible as they may be, do not apply uniformly and that there are situations in which deviations from rational behavior are systematic and substantial.

An immediate practical consequence of these findings is that one has to be keenly aware of the environment in which choices are made; the details of the buying situation matter in terms of customers' responses. How prices are presented, what "reference point" the customer perceives, the framing of the choice decision, their sense of "ownership" over the product—all can potentially influence their responses. While many of the tactics used to influence these factors lie in the domain of general marketing, the message that the choice environment can heavily influence purchase behavior is an important one for pricing modelers to heed. Moreover, such behavioral theories of demand are working their way into the operational pricing research literature, and it is likely that such ideas will influence analytical pricing practice more directly in the years ahead.

4.6 Strategic Behavior Models

Another important customer behavior that has important consequences for tactical pricing is the self-interested behavior of customers. Such considerations lead to the idea of *strategic customer models*, in which customers are viewed as utility maximizers who respond to a seller's actions and adjust their timing and channel of purchase in order to maximize their gains. Aviv and Vulcano [THIS VOLUME] provide a comprehensive overview of strategic behavior models; here, we briefly summarize the main ideas.

Building models of this sort requires expanding the customer's strategy space – that is, the range of alternatives available to them. So, for example, rather than viewing customers as arriving and making a choice decision on the spot, we model them as deciding when to arrive as well. If they anticipate that prices will be low in the future, this then creates an incentive for them to delay their purchases. Hence, our decisions on pricing or availability over time will affect the purchase strategies adopted by our consumers. From a modeling standpoint, such behavior can be considered to be a game between a firm and its customers; each strategy adopted by the firm induces a different strategy among its customers. The question then is: Which pricing strategy should we adopt to induce the most profitable equilibrium among customers? Equilibrium and game theory methods can be used to analyze such questions.

The outputs from such models can result in profoundly different pricing decisions. Take markdown pricing; a typical pricing optimization model upon seeing sluggish sales at the current price recommends a markdown. But what if demand is low early in the sales season precisely because customers anticipate the firm will markdown eventually? Does it really make sense then to lower prices, because doing so may simply reinforce their expectations and perpetuate a bad equilibrium? Perhaps a constant pricing policy, while less profitable in the short run, would train customers not to wait for markdowns. The resulting shift in their equilibrium behavior might then be more profitable in the long run.

While such considerations have intriguing implications and there has been considerable recent theoretical research on strategic consumer models (See Shen and Xu [36] for a recent survey.), to date they have yet to find their way into tactical pricing systems. There are several reasons for this. First, the area is young and most of the research work on the topic has been done only in the last decade; new models and methods are still in being developed and disseminated. Second, strategic behavior is complex to model and analyze. Most of the current research models are quite stylized – and even these highly pared-down models often require intricate equilibrium analysis to understand. Getting such models and analysis to the level of sophistication required for realistic operational pricing applications will be a significant challenge. Lastly, it is an open question how well one can identify and estimate strategic behavior in practice. It requires understanding customers' expectations of the future and their decision making process over time. This significantly greater scope in the estimation and validation problem is a potential practical limitation.

That said, strategic consumer models hold great potential to help temper the “overly myopic” nature of current tactical pricing systems, which can overweight short-term revenue gain at the expense of long-run customer reactions.

5 Estimation and Forecasting

Estimating demand models and accurately forecasting the effect of pricing actions is critical to success in operational pricing. In many ways, though, this task is a standard one and system designers can avail themselves of a vast array of tools and techniques from statistical estimation and forecasting to accomplish it. Yet there are some unique challenges in demand model estimation which we highlight here. There are many excellent references on estimation and forecasting methods, so our focus here is on the key challenges involved in estimating demand models, the high-level design choice one must make in building an estimation and forecasting system and how demand estimation is executed in practice.

In operational pricing, estimation and forecasting are typically automated, transactional, and data-driven—as opposed to qualitative (such as expert opinion) or survey-based. This is due to the sheer volume of forecasts that have to be made and the tight processing time requirements of a real-time system. These practical constraints limit the choice of methods. They also limit the types of data that can reasonably be collected and the amount of time a user can spend calibrating and verify estimates and forecasts. Certain procedures, even if they give superior estimates, may simply not be viable options in practice because they take too long to run, require data that is too expensive to collect (e.g. surveys), or require too much expert, manual effort to calibrate.

Estimation and forecasting requirements are also driven by the input requirements of the optimization modules they feed. Many optimization models use stochastic models of demand and hence require an estimate of the complete probability distribution or at least parameter estimates (e.g., means and variances) for an assumed distribution. Besides producing estimates of price response and demand volume, many other features of the demand might need to be estimated—how it evolves over time, how it varies seasonally, from which channels and segments it arrives, how it responds to a promotions vs. a regular price change—all of which are important in making good tactical pricing decisions.

Since no statistical model can incorporate all possible factors influencing demand, it is common practice to rely on analysts to monitor outside events and compensate for special events by adjusting forecasts appropriately through so-called *user influences*. In this sense, one should view estimation and forecasting in practice as a hybrid of automated, analytical inputs and human, subjective inputs.

We begin by looking at typical data sources and main design choices in estimation and forecasting. We then survey the main methodologies used, focusing again on the those that are of special importance for demand estimation. Lastly, we discuss issues involved in implementing estimation and forecasting in real-world systems.

5.1 Data Sources

Data is the life-blood of any estimation and forecasting system. Therefore, identifying which sources of data are available and how they can best be used is an important first step in approaching the problem.

The main sources of data in most pricing systems are transactional databases—for example, reservation and property management systems (PMSs) in hotels, customer relationship management (CRM) systems, enterprise resource planning (ERP) systems, and retail inventory and scanner databases. These sources may be centralized, independent entities shared by other firms in the industry (such as global distribution systems (GDSs) of the airline industry selling MIDT data), a centralized facility within a company that interfaces with several local systems (e.g. a retail chain's point-of-sale (POS) system linking all its stores), a local reservation system (a hotel PMS), or a customer-oriented database with information on individual customers and their purchase history (CRMs).

In addition to sales information, databases often store information on the controlling process itself. Examples of this kind of data include records of past prices, promotion activities and win/loss data on customized pricing offers. Inventory data is also provided by many retail POS systems, and this data is useful for correcting for stockouts and broken-assortments effects (missing color-size combinations).

Panel data, obtained from tracking purchases of a group of panelists over time, provide valuable information on cross-sectional and intertemporal purchase behavior. Such data are widely used in retail and media industries. A panel member's purchase data is also linked to promotions, availability, displays, advertising, couponing, and markdowns through the time of purchase, allowing for precise inferences on preferences and marketing influences. Many marketing research companies provide such panel-data services.

A few auxiliary data sources also play an important role in pricing systems in some industries. For instance, currency exchange-rate and tax information is necessary to keep track of revenue value for sales in different countries. In the airline industry, the schedules and possible connections (provided by firms such as the official airline guide (OAG)) are required to determine which markets are being served. In broadcasting, ratings, customer location, and demographic information is required. A causal forecasting method may take into account information on the state of the economy, employment, income and savings rates, among other factors. Information on ad-hoc events (special events) like conferences, sports events, concerts, holidays, is also crucial in improving the accuracy of forecasts.

Many retail pricing systems also use weather data, which are supplied by several independent vendors via daily automated feeds. Short-term weather forecasts guide discounting and stocking decisions. Weather data also play an important role in energy forecasting for electric power generators and distributors.

Macroeconomic data (such as gross national product (GNP) growth rates and housing starts) are rarely used in automated, tactical forecasting but frequently play a role in aggregate forecasts of factors such as competitors' costs, industry demand and market share, and broad consumer preferences. Statistics on cost of labor are published by the Bureau of Labor Statistics (BLS) in the United States in a monthly publication called *Employment and Earnings*, which provides average hourly earnings for workers by product category. BLS also provides monthly producer price indexes on raw materials.

5.2 Estimation and Forecasting Design Decisions

Before applying any analytical methods, there are a number of important structural design choices involved in developing an estimation and forecasting system. We discuss the main choices next.

5.2.1 Bottom-Up versus Top-Down Strategies

Broadly speaking, there are two main ways to construct estimates and forecasts: bottom-up and top-down. In a *bottom-up forecasting strategy*, forecasting is performed at a detailed level to generate *subforecasts*. The end forecast is then constructed by aggregating these detailed subforecasts. In a *top-down forecasting strategy*, forecasts are made at a high level of aggregation—a *superforecast*—and then the end forecast is constructed by disaggregating these superforecasts down to the level of detail required.

Which strategy is most appropriate is not always clear-cut. The choice depends on the data that are available, the outputs required, and the types of forecasts already being made. Moreover, the "right" answer in most cases is that both strategies are required, because certain phenomena can only be estimated at a low level of aggregation, while others can only be estimated at a high level of aggregation. For example, seasonality is hard to identify at low levels of aggregation because the data is often so sparse that one cannot distinguish "peak" from "off-peak" periods. In fact, aggregate phenomena such as daily or weekly seasonalities, holiday effects, or upward or downward trends in total demand are—for all practical purposes—unobservable at the highly disaggregate level; one must look at aggregate data to estimate them. At the same time, if we aggregate across products or markets to understand seasonality, we may lose information about the difference in price sensitivity between product and markets. For this reason, using hybrid combinations of bottom-up and top-down approaches is the norm in practice.

5.2.2 Levels of Aggregation

Which level of aggregation to use (whether arrived at from a top-down or bottom up approach) is another important design decision. The choice here is not purely a matter of estimation; it is also highly dependent on the parameter requirements of the optimization model. For example in retailing, store-level pricing requires store-level estimates of demand and price sensitivity for each product, whereas a model that optimizes prices that are set uniformly on a chain-wide basis does not require such detail. If household purchase data (panel data) is available or if experiments or surveys can be conducted, then one can forecast based on models of individual purchase behavior and combine these to determine an aggregate demand function. However, if only aggregate POS sales data can be obtained, we might be limited to estimating an aggregate demand function directly.

In theory one would like the most detailed demand model possible – to know the response parameters of every customer even. That way, we can forecast at as detailed a level as we like and we don't lose information by assuming groups of customers or products have the same behavior. But as one moves down to finer levels of detail, the data becomes sparser (e.g. imagine the data a grocery store has available for total sales of a given product versus the data it has on a given customer's purchase of that same product) – and sparse data increases estimation error. There is a balance then between the *specification error* (error in the form of the model) introduced from aggregating potentially dissimilar entities together versus the *estimation error* (error in the parameter estimates of a given model) introduced by the limited data available from disaggregating entities. While there is theory from model selection in statistics to help guide this trade-off (See Burnham and Anderson [6].), in practice it's often a matter of extensive trial and error to determine which level of aggregation produces the most accurate estimate of demand.

5.2.3 Parametric vs. Nonparametric Models

Estimators can be specified in one of two ways. The first is to assume a specific functional form and then estimate the parameters of this functional form. This approach is called *parametric estimation.* Alternatively, distributions or functions can be estimated directly based on observed historical data, without assuming any *a priori* functional form. This approach is called *nonparametric estimation.* Choosing between a parametric or nonparametric approach to estimation and forecasting is a key design decision.

While nonparametric methods are in a sense more general, they are not necessarily a better choice. Nonparametric estimates suffer from two serious drawbacks: First, because they do not use a functional form to "fill in" for missing values, they often require much more data than are available in many applications to obtain reasonable estimates of a distribution or demand function. Second, even with sufficient data, nonparametric estimates may not be as good at predicting the future, even if they fit the historical data well. Parametric models are better able to "smooth out" the noise inherent in raw data, which often results in a more robust forecast. There are intermediate approaches too. Neural networks are sometimes viewed as *semiparametric* methods, in that they assume a parametric form but it is a general and high-dimensional one that, as the size of the network increases, can provide a close approximation to a non-parametric model.

Parametric methods usually are more modest in their data requirements, have the advantage of providing estimates of demand that extend beyond the range of the observed data (allow for extrapolation), and are generally more robust to errors and noise in the data. The disadvantage of parametric techniques is that some properties of the distribution must be assumed—for example, that it is symmetric about the mean, has certain coefficients of variation, or has certain *tail behavior* (the characteristics of the demand distribution for extreme values of demand). Thus, parametric methods can suffer in terms of overall forecasting accuracy if the actual demand distribution deviates significantly from these assumptions (called *specification* *errors*).

5.3 Estimation Methods

Here we briefly survey methods for estimation and discuss some of the theoretical and practical issues that arise.

5.3.1 Estimators and Their Properties

An estimator represents, in essence, a formalized "guess" about the parameters of the underlying distribution from which a sample (the observed data) is drawn. Estimators can take on many forms and can be based on different criteria for a "best" guess. We focus on parametric estimation.

For a parametric estimator, we assume that the underlying distribution of demand, Z, is of the form

$$P(Z \leq z|\boldsymbol{\beta}, \mathbf{y}) \quad = \quad F(z, \boldsymbol{\beta}, \mathbf{y}), \tag{22}$$

where $\mathbf{y} = (y_1, \ldots, y_M)$ is a vector of M observed explanatory variables which includes, of course, prices (own and competitors) - but can also include other explanatory variables such as time, indicators of holiday events, lagged observations of Z itself, and so on. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$ is a M-dimensional vector of parameters that must be estimated from data. Z is often referred to as the *dependent* variable and the vector $\mathbf{y}$ as the *independent* variables. The density function of demand (if it exists) is denoted $f(z|\boldsymbol{\beta}, \mathbf{y}) = \frac{d}{dz} F(z, \boldsymbol{\beta}, \mathbf{y})$. For ease of exposition, we assume that the dimensions of $\boldsymbol{\beta}$ and $\mathbf{y}$ are the same, though this is not necessary. For example, this function may be one of the aggregate demand functions discussed above or a random utility model of individual purchase outcomes.

Alternatively we can express the relationship between the demand and the explanatory variables as consisting of two parts: a systematic (deterministic) component, $\zeta(\boldsymbol{\beta}, \mathbf{y}_k)$ (also called a *point estimate*), and a zero-mean error term, ξ_k, so that:

$$Z_k = \zeta(\boldsymbol{\beta}, \mathbf{y}_k) + \xi_k. \tag{23}$$

Assume we have a sequence of N independent observations $z_1, \ldots, z_N$, with values for the explanatory variables represented by vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$. The estimation problem, then, is to determine the unknown parameters $\boldsymbol{\beta}$ using only the sample of the N observations and the values of the explanatory variables corresponding to each observation. The following is a simple example:

As an example, consider the linear demand function with additive noise:

$$Z = \boldsymbol{\beta}^\top \mathbf{y} + \xi, \tag{24}$$

where ξ is and i.i.d. $N(0, \sigma^2)$ random noise term, independent of the explanatory variables $\mathbf{y}$. The distribution of Z in terms of (22) is then

$$F_Z(z|\boldsymbol{\beta}, \mathbf{y}_k) = P(Z \leq z|\boldsymbol{\beta}, \mathbf{y}) = \Phi(\frac{z - \boldsymbol{\beta}^\top \mathbf{y}}{\sigma}),$$

where $\Phi(\cdot)$ is the standard normal distribution.

5.3.2 Properties of Estimators

If the set of N observations, $\mathbf{z}_N = (z_1, \ldots, z_N)$, are considered independent realizations of Z, then an estimator based on these observations is a function of the random variables, $\hat{\boldsymbol{\beta}}(\mathbf{z}_N)$, and is therefore itself a random variable. What properties would we like this (random) estimator to have?

Unbiasedness For one, it would be desirable if the expected value of the estimator equaled the actual value of the parameters—that is, if

$$E[\hat{\boldsymbol{\beta}}(\mathbf{z}_N)] = \boldsymbol{\beta}.$$

If this property holds, the estimator is said to be an *unbiased estimator*; otherwise, it is a *biased estimator*. The estimator of the m^{th} parameter, $\hat{\beta}_m$, is said to have a *positive bias* if its expected value exceeds β_m, and a *negative bias* if its expected value is less than β_m.

If the estimator is unbiased only for large samples of data—that is, it satisfies

$$\lim_{N \to \infty} E[\hat{\boldsymbol{\beta}}(\mathbf{Z}_N)] = \boldsymbol{\beta}$$

—then it is called an *asymptotically unbiased estimator*.

Efficiency An estimator $\hat{\boldsymbol{\beta}}(\mathbf{Z})$ is said to an *efficient estimator* if it is unbiased and the random variable $\hat{\boldsymbol{\beta}}(\mathbf{Z})$ has the smallest variance among all unbiased estimators. Efficiency is desirable because it implies the variability of the estimator is as low as possible given the available data. The Cramer-Rao bound[12] provides a lower bound on the variance of *any* estimator, so if an estimator achieves the Cramer-Rao bound, then we are guaranteed that it is efficient. An estimator can be inefficient for a finite sample but *asymptotically efficient* if it achieves the Cramer-Rao bound when the sample size is large.

[12] See DeGroot [11], pp. 420–430 for a discussion of the Cramer-Rao bound.

Consistency An estimator is said to be *consistent* if for any $\delta > 0$,

$$\lim_{N\to\infty} P(|\hat{\boldsymbol{\beta}}(\mathbf{Z}_N) - \boldsymbol{\beta}| < \delta) = 1.$$

That is, if it converges in probability to the true value $\boldsymbol{\beta}$ as the sample size increases. Consistency assures us that with sufficiently large samples of data, the value of $\boldsymbol{\beta}$ can be estimated arbitrarily accurately.

Ideally, we would like our estimators to be unbiased, efficient, and consistent, but this is not always possible.

5.3.3 Minimum Square Error (MSE) and Regression Estimators

One class of estimators is based on the *minimum mean-square error* (MSE) criterion. MSE estimators are most naturally suited to the case where the forecast quantity has an additive noise term as in (23). Given a sequence of observations $z_1, \ldots, z_N$ and associated vectors of explanatory variable values $\mathbf{y}_1, \ldots, \mathbf{y}_N$, the MSE estimate of the vector $\boldsymbol{\beta}$ is the solution to

$$\min_{\boldsymbol{\beta}} \sum_{k=1}^{N} [z_k - \zeta(\boldsymbol{\beta}, \mathbf{y}_k)]^2, \tag{25}$$

where the point estimate $\zeta(\boldsymbol{\beta}, \mathbf{y}_k)$ is as defined in (23). The minimization problem (25) can be solved using standard nonlinear optimization methods such as conjugate-gradient or quasi-Newton. If the point estimate is linear so that

$$Z = \boldsymbol{\beta}^\top \mathbf{y} + \xi.$$

and the error term ξ is a normal random variable that is i.i.d. with zero mean and constant variance for all observations, then the MSE estimate is known as the *ordinary least-squares (OLS) estimators*—or *linear-regression estimators*. Standard linear regression packages can be used to estimate such models. Because of its simplicity and efficiency, regression is widely used in price-based management for estimating price sensitivity, market shares, and the effects of various marketing variables (such as displays and promotions) on demand.

5.3.4 Maximum-Likelihood (ML) Estimators

While regression is based on the least-squares criterion, *maximum-likelihood (ML) estimators* are based on finding the parameters that maximize the "likelihood" of observing the sample data, where *likelihood* is defined as the probability of the observations occurring. More precisely, given a probability-density function f_Z of the process generating a sample of data $Z_k,\ k = 1, \ldots, N$, which is a function of a vector of parameters $\boldsymbol{\beta}$ and the observations of the explanatory variables, $\mathbf{y}_k$, the likelihood of observing value z_k as the k^{th} observation is given by the density $f_Z(z_k|\boldsymbol{\beta}, \mathbf{y}_k)$ (or by the probability mass function if the demand distribution is discrete). The likelihood of observing the N observations $(z_1, \mathbf{y}_1), \ldots, (z_N, \mathbf{y}_N)$ is then

$$\mathcal{L} = \prod_{k=1}^{N} f_Z(z_k|\boldsymbol{\beta}, \mathbf{y}_k). \tag{26}$$

The ML estimation problem is to find a $\boldsymbol{\beta}$ that maximizes this likelihood $\mathcal{L}$. It is more convenient to maximize the log-likelihood, $\ln \mathcal{L}$, because this converts the product of function in (26) to a sum of functions. Since the log function is strictly increasing, maximizing the log-likelihood is equivalent to maximizing the likelihood. This gives the ML problem:

$$\max_{\boldsymbol{\beta}} \sum_{k=1}^{N} \ln f_Z(z_k|\boldsymbol{\beta}, \mathbf{y}_k).$$

In special cases, this problem can be solved in closed form. Otherwise, if the density $f_Z(\cdot)$ (or probability mass function in the discrete case) is a differentiable function of the parameters $\boldsymbol{\beta}$, then gradient-based optimization methods can be used to solve it numerically.

ML estimators have good statistical properties under very general conditions; they can be shown to be consistent, asymptotically normal, and asymptotically efficient, achieving the Cramer-Rao lower bound on the variance of estimators for large sample sizes.

As an example, consider estimating the parameters of the MNL choice model described above. The data consists of a set of N customers and their choices from a finite set S of alternatives. Associated with each alternative j is a vector $\mathbf{y}_j$ of explanatory variables (assume for simplicity there are no customer-specific characteristics). The probability that a customer selects alternative i is then given by

$$P_i(S) = \frac{e^{\boldsymbol{\beta}^\top \mathbf{y}_i}}{\sum_{j \in S} e^{\boldsymbol{\beta}^\top \mathbf{y}_j} + 1}, \tag{27}$$

where $\boldsymbol{\beta}$ is a vector of (unknown) parameters. Let $c(k)$ be the choice made by customer k. The likelihood function is then

$$\mathcal{L} = \prod_{k=1}^{N} \left[\frac{e^{\boldsymbol{\beta}^\top \mathbf{y}_{c(k)}}}{\sum_{j \in S} e^{\boldsymbol{\beta}^\top \mathbf{y}_j} + 1} \right],$$

and the maximum-likelihood estimate $\hat{\boldsymbol{\beta}}$ is then determined by solving

$$\max_{\boldsymbol{\beta}} \; \ln \mathcal{L}. \tag{28}$$

While this maximum-likelihood problem cannot be solved in closed form, it has good computational properties. Namely, there are closed-form expressions for all first and second partial derivatives of the log-likelihood function, and it is jointly concave in most cases (McFadden [31]; Hausman and McFadden [18]). The ML estimator has also proven to be robust in practice.

5.3.5 Method of Moments and Quantile Estimators

While MSE and ML are the most prevalent estimators in practice, several other estimators are also used as well. Two common ones are *method of moments* and *quantile estimators*.

In the method of moments, one equates moments of the theoretical distribution to their equivalent empirical averages in the observed data. This yields a system of equations that can be solved to estimate the unknown parameters $\boldsymbol{\beta}$.

Alternatively, we can use quantile estimates based on the empirical distribution to estimate the parameters $\boldsymbol{\beta}$ of a distribution. For example, we might estimate the mean of a normal distribution by noting that as the normal distribution is symmetric, the mean and median are the same. Hence, we can estimate the mean by computing the median of a sequence of N observations. More generally, one can compute a number of quantiles of a data set and equate these to the theoretical quantiles of the parametric distribution. In general, if m parameters need to estimated, m different quantiles are needed to produce m equations in m unknowns (for a normal distribution, for example, one could equate the 0.25 and 0.75 quantiles of the data to the theoretical values to get two equations for the mean and variance). Quantile estimation techniques are sometimes preferred as they tend to be less sensitive to outlier data than are MSE and ML estimators.

5.4 Endogeneity, Heterogeneity, and Competition

In this section, we focus on a few estimation problems that are of particular importance for pricing applications—endogeneity, heterogeneity, and competition.

5.4.1 Endogeneity

The model (23) is said to suffer from endogeneity if the error term ξ is correlated with one of the explanatory variables in $\mathbf{y}$. This is a common problem in pricing practice, both in aggregate-demand function estimation and in disaggregate, discrete-choice model estimation. For example, products may have some unobservable or unmeasurable features—quality, style, reputation—and the selling firm typically prices its products accordingly. So if there are two firms in the market with similar products and one has higher nonquantifiable quality, we may observe that the firm with the higher-quality product has both a larger market share and a higher price. A naive estimate based on market shares that ignores the unobserved quality characteristics would lead to the odd conclusions that higher price leads to higher market share. Or take the case of a model that ignores seasonality; demand is higher in a peak season and firms typically price high accordingly. Without accounting for the resulting correlation between the price and this unobserved seasonal variation in demand, we may again reach the false conclusion that high prices lead to high demand. Lastly, consider a car dealer who sizes up a customer based on the way they dress, act and the information they reveal about their profession. Based on such information, the dealer may quote higher prices to customers who have higher willingness to pay, and hence only looking at transaction data we may conclude that higher prices "produce" higher probabilities of sale.

Econometricians call this problem *endogeneity* or *simultaneity*. The technical definition is that the random-error term in (23) is correlated with one of the explanatory variables, $E[\mathbf{Y}^{\top}\boldsymbol{\xi}] \neq 0$, or equivalently (in the case of linear regression) these vectors are not orthogonal. So while $\boldsymbol{\xi}$ is supposed to represent all unobservable customer and product characteristics that influence demand for a given set of explanatory variables $(Z|\mathbf{y})$, some of the explanatory variables $\mathbf{y}$ also contain information on the unobservable attributes through their correlation with $\boldsymbol{\xi}$. Such effects are a common problem in price elasticity estimation. Indeed, a recent meta-analysis of elasticity estimates by Bijmolt et al. [5] showed that accounting for endogeneity was the single strongest determinant of price elasticity differences among a wide set of elasticity studies.

Econometric techniques for correcting for endogeneity fall under a class of methods called *instrumental-variables (IV) techniques*, attributed to Reiersøl [35] and Geary [14]. Two-stage and three-stage least-squares methods (2SLS and 3SLS) are some of the popular IV techniques. Instrumental variables are exogenous variables that are correlated with an explanatory variable but are uncorrelated with the error term ξ. If there are such IVs, we can use them to "remove" the problematic correlation between the independent variables $\mathbf{y}$ and $\boldsymbol{\xi}$.

For example, one often encounters endogeneity when estimating discrete-choice demand models such as the MNL model from aggregate data (prices correlated with unobservable product characteristics). However, the problem is hard to correct because the aggregate demand is a nonlinear function of the utilities of each product and the endogeneity is present in the equation for the utilities. So using any IV technique for correcting for endogeneity becomes computationally challenging, as pointed out by Berry [2]. Berry [2] and Berry, Levinsohn, and Pakes [3] recommend that for the case of discrete-choice models in an oligopoly setting, one use measures of the firm's costs and the attributes of the products of the other firms as IVs. See also Besanko, Gupta, and Jain [4] for estimating a logit model in the presence of endogeneity due to competition.

5.4.2 Heterogeneity

We have already examined a few models of heterogeneity—namely, the finite-mixture logit model and the random-coefficients discrete-choice model. Here, we discuss how to estimate these models.

Estimation of the finite-mixture logit model is relatively straightforward. First, we must determine the number of segments. If there is no *a priori* knowledge of the number, we iterate the estimation procedure, increasing or decreasing the number of segments in each round, using suitable model-selection criteria to decide on the optimal number of segments. For a given number of segments L, we find the parameters that maximize the log-likelihood function. Estimating the random-coefficient logit, likewise, can be done using maximum-likelihood methods, though it is more difficult in general than the standard multinomial logit.

One of the problems dealing with unobservable heterogeneity in the population is that we often have to assume a distribution of heterogeneity without having much evidence as to its specification. Many times, a distribution is chosen for analytical or computational convenience. Unfortunately, a situation can arise where two radically different distributions of heterogeneity equally support the same aggregate demand observations.

Heckman and Singer [19], illustrate this overparameterization with the following example: Consider an aggregate-demand function based on a heterogeneity parameter θ. The variance on the distribution of θ represents the degree of heterogeneity. Let the demand for a particular value of θ be given by the distribution

$$G_1(z|\theta) = 1 - e^{-z\theta}, \quad z \geq 0,\ \theta > 0,$$

and let θ be equal to a constant η with probability 1 (essentially saying the population is homogeneous). The aggregate-demand distribution then is $F_1(z) = 1 - e^{-z\eta}$.

Consider another possible specification where

$$G_2(z|\theta) = 1 - \int_{z(2\theta)^{-0.5}}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-w^2/2} dw, \quad z \geq 0$$

and the distribution of θ given by $\eta^2 e^{-\eta^3\theta}$. This also turns out to lead to an aggregate-demand distribution given by $1 - e^{-z\eta}$. So based only on aggregate demand data, it is impossible to identify which specification is correct.

Nonparametric methods avoid the problem of having to specify a distribution, and Jain, Vilcassim, and Chintagunta [24] follow this strategy. Assume that the coefficients of the MNL model $\boldsymbol{\beta}$ in (27) are randomly drawn from a discrete multivariate probability distribution $G(\boldsymbol{\Theta})$. That is, the k^{th} customer is assumed to make his choice using $\tilde{\boldsymbol{\beta}}_k$, whose components are drawn from $G(\boldsymbol{\Theta})$. $G(\cdot)$ is considered a discrete distribution with support vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L$. They estimate the number of support vectors L, the location of the support vectors, and the probability mass $\boldsymbol{\theta}_i$ associated with the i^{th} support vector from observed data.

5.4.3 Competition

Accounting for the effects of competition when estimating demand models is another common and important challenge. The most direct approach - and one frequently used in practice - is simply to use competitors' prices as explanatory variables in a standard regression or discrete choice model of demand. The potential problem with this approach is that it implicitly assumes that competitors will not react to your price change - or alternatively, when a firm makes a price change, analysts have to manually estimate the reaction of competitors and input this set of prices into an industry model to determine demand.

Another strategy in such cases is to assume a model of competition between the firms, derive the equilibrium conditions implied by this model, and then estimate the parameters subject to these equilibrium conditions. We illustrate this approach with an example: Assume a homogeneous population of customers who choose among n products according to the MNL model. Then the theoretical share of product j is given as in Section 3.4,

$$P_j \quad = \quad \frac{e^{\boldsymbol{\beta}^\top \mathbf{y}_j}}{\sum_{i=1}^{n} e^{\boldsymbol{\beta}^\top \mathbf{y}_i}}, \tag{29}$$

where price is one of the explanatory variables in $\mathbf{y}_j$. One way to estimate the parameters $\boldsymbol{\beta}$ is by equating the observed market share to the theoretical equilibrium prediction. It is convenient to take logs in doing this, which yields the following system of equations relating market shares to choice behavior:

$$\ln P_j = \boldsymbol{\beta}^\top \mathbf{y}_j - \ln(\sum_{i=1}^{n} e^{\boldsymbol{\beta}^\top \mathbf{y}_i}), \quad j = 1, \ldots, n. \tag{30}$$

Next assume that prices are formed by a Bertrand style competition in prices. Let c_j be the constant marginal cost of production for product j. The profit function for product j is given by

$$V_j(p_j) \quad = \quad (p_j - c_j) N P_j, \tag{31}$$

where N is the size of the population. Let β_p be the coefficient of price in (29). Differentiating (31) with respect to p_j and setting it to zero, we get the first-order equilibrium conditions,

$$(p_j - c_j)\beta_p P_j (1 - P_j) + P_j \quad = \quad 0, \quad j = 1, \ldots, n. \tag{32}$$

The vector of parameters $\boldsymbol{\beta}$ is then estimated by attempting to fit a solution to (30) and (32) simultaneously. This can be done using, say, nonlinear least-squares estimation.

6 Conclusions

A model of demand is the heart and sole of an analytical pricing. And constructing good ones is a complex task – requiring data sources, information technology for collecting and storing data, statistical estimations models and algorithms to process and analyze these data and infrastructure for deploying model outputs - all of this is required to turn raw data into actionable market decisions.

Modern pricing systems are arguably among the most advanced applications of business analytics today. Only the most sophisticated models used in financial markets (e.g. portfolio optimization, options pricing, etc.) and those in certain direct marketing settings (e.g. catalog retailing, credit cards, etc.) rival the complexity and sophistication of today's state-of-the-art pricing systems. But in absolute terms, there is still a long way to go. Pricing systems today use only a fraction of the relevant data that today's information-enabled business environment makes possible. And currently only the basic factors affecting demand are estimated from these data. Much more is possible. The behavioral economic and strategic consumer models emerging in the research literature are promising in terms of expanding the scope and sophistication of demand models, but more development will be needed to bring these ideas into practice.

More complex models of behavior will in turn drive the need for increasingly sophisticated optimization and estimation methods. There may come a time where the phenomenon and problems encountered become so complex that modeling itself becomes the main obstacle. In this case, it may be necessary to resort to more computational approaches like agent-based modeling and simulation to make progress. There will certainly be no shortage of interesting challenges in the years ahead.

7 Further Reading

The books of Phillips [34] and Talluri and van Ryzin [37] provide comprehensive treatment of demand models and their application to tactical pricing and revenue management. The book by Nagle [33] provides a good general management overview of pricing decisions. Two excellent and comprehensive references on discrete choice models are Ben-Akiva and Lerman [1] and Train [38]. Both include extensive treatment of choice model estimation as well. Kök et al. [27] provide a review of applications of demand models to retail assortment optimization. Shen and Xu [36] provide a survey of recent research on strategic customer models. See also Ho and Xu [21] and Elmaghraby and Keskinocak [13].

For details on estimating price-response functions and market-share models, see the following marketing science text books: Eliashberg and Lilien [12], Wedel and Kamakura [42], Hanssens, Parsons, and Schultz [17], Cooper and

Nakanishi [8], Dasgupta, Dispensa, and Ghose [9], Hruska [23], West, Brockett, and Golden [43], Hill et al. [20], Zhang [44], and Lee et al. [29]. See Berry, Levinshohn, and Pakes [3], Berry [2], Besanko, Gupta, and Jain [4] and Chintagunta, Kadiyali, and Vilcassim [7] for estimation in a competitive markets. The problem of endogenity in estimation has received much recent attention in the marketing science literature spurred by the paper of Berry [2]. See also Chintagunta, Kadiyali and Vilcassim [7] and Villas-Boas and Winer [39].

References

[1] M. Ben-Akiva and S. Lerman. *Discrete-Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA, 1985.

[2] S. Berry. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25:242–262, 1994.

[3] S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63:841–890, July 1995.

[4] D. Besanko, S. Gupta, and D. Jain. Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science*, 44:1533–1547, 1998.

[5] van Heerde H. J. Bijmolt, H. A. and R. G. M. Pieters. New empirical generalizations of the determinants of price elasticity. *Journal of Marketing Research*, 42:141–156, 2005.

[6] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical-Theoretic Approach (2nd ed.)*. Springer-Verlag, New York, 2002.

[7] P. K. Chintagunta, V. Kadiyali, and N. J. Vilcassim. Endogeneity and simultaneity in competitive pricing and advertising: A logit demand analysis. Technical report, Graduate School of Business, University of Chicago, Chicago, IL, January 1999.

[8] L. G. Cooper and M. Nakanishi. *Market-share analysis.* Kluwer, Norwell, MA, 1988.

[9] C. G. Dasgupta, G. S. Dispensa, and S. Ghose. Comparing the predictive performance of a neural network model with some traditional

market response models. *International Journal of Forecasting*, 10:235–244, 1994.

[10] G. Debreu. Review of R. D. Luce, individual choice behavior: A theoretical analysis'. *American Economic Review*, 50:186–188, 1960.

[11] M. H. DeGroot. *Probability and Statistics*. Addison Wesley, Reading, MA, 2nd edition, 1985.

[12] J. Eliashberg and G. L. Lilien, editors. *Management Science in Marketing*. North Holland, Amsterdam, The Netherlands, 1991.

[13] W. J. Elmaghraby and P. Keskinocak. Dynamic pricing: Research overview, current practices, and future directions. *Management Science*, 2003. (To appear).

[14] R. Geary. Determination of linear relations between systematic parts of variables with errors of observation the variance of which is unknown. *Econometrica*, 17:30–58, 1949.

[15] P. M. Guadagni and J. D. C. Little. A logit model of brand choice calibrated on scanner data. *Marketing Science*, 3:203–238, 1983.

[16] J. D. Gwartney and R. L. Stroup. *Economics: Private and Public Choice*. Harcourt, Stamford, CT, 8th edition, 1997.

[17] D. M. Hanssens, L. J. Parsons, and R. L. Schultz. *Market Response Models: Econometric And Time Series Analysis*. Kluwer, Norwell, MA, 1990.

[18] J. Hausman and D. McFadden. A specification test for the multinomial logit model. *Econometrica*, 52:1219–1240, 1984.

[19] J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52:271–320, 1984.

[20] T. Hill, M. O'Connor, and W. Remus. Neural network models for time series forecasting. *Management Science*, 42:1082–1091, 1994.

[21] T-H. Ho and X. Su. Strategic pricing response and optimization in operations management. In V. R. Rao, editor, *Handbook of Pricing Research in Marketing*. Edward Elgar Publishing Company, Cheltenham, UK, 2009.

[22] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1994.

[23] H. Hruska. Determining market response functions by neural network modeling: A comparison to econometric techniques. *European Journal of Operational Research*, 66:27–35, 1993.

[24] D. Jain, N. J. Vilcassim, and P. K. Chintagunta. A random-coefficients logit brand-choice model applied to panel data. *Journal of Business and Economic Studies*, 12:317–328, 1994.

[25] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgement Under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, UK, 1982.

[26] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–292, 1979.

[27] Fisher M. Kok, G. and R. Vaidyanathan. Assortment planning: Review of literature and industry practice. In N. Agrawal and S. Smith, editors, *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, pages 99–154. Springer, New York, New York, 2009.

[28] D. M. Kreps. *Notes on the Theory of Choice.* Westview Press, London, UK, 1988.

[29] K. Lee, T. Choi, C. Ku, and J. Park. Neural network architectures for short term load forecasting. In *Proceedings of the IEEE International Conference on Neural Networks*, Orlando, FL, 1994. IEEE ICNN.

[30] R. Luce. *Individual Choice Behavior: A Theoretical Analysis.* Wiley, New York, NY, 1959.

[31] D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, NY, 1974.

[32] D. McFadden and K. Train. Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15:447–470, 2000.

[33] N. T. Nagle and R. K. Holden. *The Strategy and Tactics of Pricing: A Guide to Growing More Profitably.* Pearson Education, Upper Saddle River, Nwe Jersey, 4th edition, 2006.

[34] R. L. Phillips. *Pricing and Revenue Optimization.* Stanford University Press, Stanford, California, 2005.

[35] O. Reiersøl. Confluence analysis by means of instrumental sets of variables. *Arkiv for Mathematik, Astronomi och Fysik*, 32, 1945.

[36] Z.J.M. Shen and X. Su. Customer behavior modeling in revenue management and auctions: A review and new research directions. *Production and Operations Management*, 16:713–728, 2007.

[37] K. T. Talluri and G. J. van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, January 2004.

[38] K. E. Train. *Discrete Choice Methods with Simulation.* Cambridge University Press, 2006.

[39] J. M. Villas-Boas and R. S. Winer. Endogeneity in brand-choice models. *Management Science*, 45:1324–1338, 1999.

[40] X. Vives. *Oligopoly Pricing: Old Ideas and New Tools.* MIT Press, Cambridge, MA, 2001.

[41] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, NJ, 1944.

[42] M. Wedel and W. Kamakura. *Market segmentation: Conceptual and Methodological Foundations.* Kluwer, Norwell, MA, 2nd edition, 2000.

[43] P. M. West, P. L. Brockett, and L. L. Golden. A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, 16:370–391, 1997.

[44] G. Zhang, B. E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14:35–62, 1998.

[45] S. Ziya, H. Ayhan, and R. D. Foley. On assumptions ensuring a strictly unimodal revenue function. Technical report, 2004.