

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

9-2008

Choice Models in Marketing: Economic Assumptions, Challenges and Trends

CHANDUKALA, Sandeep R.

Singapore Management University, sandeepc@smu.edu.sg

Jaehwan KIM

Korea University - Korea

Thomas OTTER

Johann Wolfgang Goethe Universitat Frankfurt am Main

Peter E. ROSSI

University of Chicago

Greg M. ALLENBY

Ohio State University - Main Campus

DOI: <https://doi.org/10.1561/1700000008>

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Marketing Commons](#)

Citation

CHANDUKALA, Sandeep R.; KIM, Jaehwan; OTTER, Thomas; ROSSI, Peter E.; and ALLENBY, Greg M.. Choice Models in Marketing: Economic Assumptions, Challenges and Trends. (2008). *Foundations and Trends in Marketing*. 2, (2), 97-184. Research Collection Lee Kong Chian School Of Business.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/4801

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Choice Models in Marketing: Economic Assumptions, Challenges and Trends

Sandeep R. Chandukala¹, Jaehwan Kim²,
Thomas Otter³, Peter E. Rossi⁴
and Greg M. Allenby⁵

¹ *Kelley School of Business, Indiana University, Bloomington, IN 47405-1701, USA, sarchand@indiana.edu*

² *Korea University Business School, Korea University, Korea, jbayes@korea.ac.kr*

³ *Department of Marketing, Johann Wolfgang Goethe-Universität Frankfurt, Germany, otter@marketing.uni-frankfurt.de*

⁴ *Graduate School of Business, University of Chicago, USA, peter.rossi@chicagogsb.edu*

⁵ *Fisher College of Business, Ohio State University, USA, allenby_1@fisher.osu.edu*

Abstract

Direct utility models of consumer choice are reviewed and developed for understanding consumer preferences. We begin with a review of statistical models of choice, posing a series of modeling challenges that are resolved by considering economic foundations based on constrained utility maximization. Direct utility models differ from other choice models by directly modeling the consumer utility function used to derive the likelihood of the data through Kuhn-Tucker conditions. Recent advances in Bayesian estimation make the estimation

of these models computationally feasible, offering advantages in model interpretation over models based on indirect utility, and descriptive models that tend to be highly parameterized. Future trends are discussed in terms of the antecedents and enhancements of utility function specification.

1

Introduction and Scope

Understanding and measuring the effects of consumer choice is one of the richest and most challenging aspects of research in marketing. Choice comes in many varieties and forms. It can be discrete in the sense of the selection of just one item, or it can be continuous when multiple items are purchased or selected. Choice can reflect careful deliberation, habit, or a consumer's spontaneous reactions to marketing variables. It need not always result in purchases in the marketplace, or be driven by standard concepts of utility. It can represent trade-offs that may or may not be continuous or compensatory. Most interestingly, it relates to all marketing control variables (the 4 P's), as these variables enter into the decision-making process.

In this issue of "Foundations and Trends in Marketing" we examine recent developments in the modeling of choice for marketing. Choice in marketing differs from other domains in that the choice context is typically very complex, and researchers' desire knowledge of the variables that ultimately lead to demand in marketplace. The marketing choice context is characterized by many choice alternatives. Moreover, the number of attributes and features characterizing choice alternative

is often large. Identifying the variables that drive choice is challenging because consumers are heterogeneous in their use of these variables.

Researchers in marketing are also interested in understanding processes that drive preference. It is often not possible to assume the existence of a well-defined preference ordering for all product attributes and brands, and the use of simple descriptive models can mask important variables, such as the “must haves” for a product. Marketing’s role within an organization is to guide management in what to offer in the marketplace, which can be incompatible with the assumption that a preference structure already exists.

As consumers encode, process, and react to marketplace stimuli, numerous opportunities exist for identifying relevant variables, and the means by which these variables combine to form aspects of consideration, evaluation, and choice. Advances in statistical computing and the development of new hierarchical Bayes models have enabled researchers in marketing to make significant inroads to quantifying aspects of choice. These inroads, however, are merely initial steps along a path to understand and characterize how consumers make choice decisions.

The aim of this issue is to lay out the foundations of choice models and discuss recent advances. We focus on aspects of choice that are, and can be quantitatively modeled. Moreover, we only consider models that can be directly related to a process of constrained utility maximization. Thus, we discuss a portion of a large stream of research currently being developed by both quantitative and qualitative researchers in marketing. Our hope is that by reviewing the basics of choice modeling, and pointing to new developments, we can provide a platform for future research.

Marketing models of choice have undergone many transformations over the last 20 years, and the advent to hierarchical Bayes models indicate that simple, theoretically grounded models work well when applied to understanding individual choices. Thus, we use economic theory to provide the foundation from which future trends are discussed. We begin our discussion with descriptive models of choice that raises a number of debatable issues for model improvement. We then look to economic theory as a basis for guiding model development. Economic theory assumes the existence of preference orderings for which

utility can be parameterized and used to understand aspects of choice. This theory, however, is somewhat silent on how utility arises, or is constructed.

Utility construction is critical to the marketing discipline because marketing's role is to provide guidance to firms on offerings that are responsive to the needs of individuals, and to provide specifics as to how best to sell these goods. As a result, researchers in marketing have an expanded domain of study beyond traditional economics. We believe that future trends of choice models comprise elements that precede, and are implicated by, formal economic models. We briefly discuss some of these interesting areas of research.

2

Basic Assumptions

The dependent variable in all marketing models is ultimately some aspect of consumer behavior, and this behavior usually involves some form of discreteness. The most obvious example of this is choice between near-perfect substitutes, e.g., from among different brands within a product category. Consumers typically choose just one of the alternatives, and exhibit zero demand for the remainder. But, even when goods are not near-perfect substitutes, such as decisions that span more than one product category, or involve some sort of volume decision, consumer demand data can be characterized with the number zero as the most frequent response, the number one observed to be the next most frequent response, and so on. Thus, a major challenge in modeling choice is answering the question of how best to deal with zero's present in the data.

There are many statistical models that can be immediately employed to deal with zero responses. Examples include Bernoulli, binomial, multinomial, and Poisson distributions. The Bernoulli model can be expressed as

$$\pi(x|p) = p^x(1 - p)^{1-x}, \quad (2.1)$$

where p denotes the probability of observing the outcome x equal to one, and $(1 - p)$ is the probability of observing x equal to zero. We will refer to Equation (2.1) as the likelihood of the datum (x). When we observe x equal to one, the likelihood function yields the value p , and when x equals zero, the likelihood is equal to $(1 - p)$.

The algebraic form of Equation (2.1) serves as the basis of many choice models in marketing. If we use x_t to refer to a set of choices from a consumer at different points in time (t), each thought of as being generated from Equation (2.1), then the likelihood of the vector x can be written as (absent the normalizing constant):

$$\pi(x' = (x_1, x_2, \dots, x_T) | p) \propto \prod_t p^{x_t} (1 - p)^{1 - x_t}. \quad (2.2)$$

Equation (2.2) implies much about choice. It implies (i) that choices over time are independently distributed given, or conditional on, p . That is, consumers approach the choice task as if there is no “history.” Equation (2.2) also implies that (ii) the choice probability is constant over time periods. Finally, Equation (2.2) implies that demand is (iii) represented by either a zero (i.e., $x = 0$) or a one, indicating nonzero consumption.

Much of the historical choice literature in marketing over the last 30 years is aimed at addressing the three properties of Equation (2.2), i.e., conditional independence, constant choice probabilities, and constant quantity. Each of these properties is clearly violated in most marketing contexts. One can think of marketing’s role as an attempt to affect choice so that these properties are not true. Marketing’s role of guiding management involves manipulating various control variables so re-purchase probabilities are not independent, competitive effects are accurately represented, and purchase quantities are large.

A first step toward relaxing the properties of the Bernoulli likelihood in Equation (2.2) is to allow for choice among a set of relevant alternatives using a multinomial distribution. In this model, demand in each time period is represented by a vector with one vector element equal to one, and the remaining elements equal to zero. The likelihood for a single observation can be written as

$$\pi(x | p) \propto p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad (2.3)$$

where the probabilities are all positive and sum to one. The likelihood for a set of observations, $\{x_t, t = 1, \dots, T\}$ is an immediate generalization of Equation (2.2):

$$\pi(x' = (x_1, x_2, \dots, x_T) | p) \propto \prod_t p_1^{x_{1t}} p_2^{x_{2t}} \dots p_k^{x_{kt}}. \quad (2.4)$$

The use of a multinomial response model immediately raises the question of what items to include in the choice set. This issue can be included as a fourth point (iv) in addition to the three assumptions listed above (i)–(iii) as possible research problem areas that require elaboration for Equation (2.3) to be applied in a marketing context. We will see that much of the marketing literature on choice modeling originates by questioning basic and implied assumptions of such standard models.

An easy way to generalize the choice model represented in Equation (2.4) for application to marketing is to make the choice probabilities, $\{p_i\}$, a function of variables such as prices. A simple way of relating prices to choice probabilities is with something known as a logit link function:

$$p_i = \frac{\exp[\beta_{0i} + \beta_p \ln \text{price}_i]}{\sum_{j=1}^K \exp[\beta_{0j} + \beta_p \ln \text{price}_j]} \quad i = 1, \dots, K. \quad (2.5)$$

The advantage of this function is that it constrains choice probabilities to be positive and sum to one. In Equation (2.5), the choice probabilities are seen to be related to a set of brand intercepts, $\{\beta_0\}$ and prices expressed in logarithmic form. As with any addition to the basic model, new questions about this specification can be proposed and tested using marketing data. Two immediate issues involve (v) why logarithmic prices?; and (vi) why just one price coefficient? As stated earlier, there exists a long history of such questions investigated in various ways throughout the marketing literature on choice modeling, some of which is discussed below. The important thing to remember is that research springs from questioning the underlying model assumptions, offering many opportunities for understanding the context in which modeling assumptions are appropriate for describing behavior.

The introduction of Equation (2.5) raises a number of other issues. The first issue is model estimation. In this survey, we focus our attention on the use of likelihood-based methods of estimation, and do not discuss moment-based methods such as GMM. The reason is due to a statistical principle first proposed by Fisher (1922) known as the likelihood principle (see also Liu et al., 2007). The likelihood principle states that all learning about model parameters such as brand intercepts $\{\beta_0\}$ and price coefficients (β_p) from the data occur through the likelihood function. Thus, the likelihood is sufficient for representing the information in the data. Moreover, the likelihood contains information from every moment of the data, while moment-based estimators only use information from specific moments. Thus, likelihood based methods are more efficient. A second reason, discussed further below, is that advances in Bayesian computation allows access to classes of choice models that are hard to estimate accurately with other methods. Bayesian analysis adheres to the likelihood principle, and therefore results in more accurate estimation for a wider class of models than other methods.

Maximum likelihood estimation of the model implied by Equations (2.4) and (2.5) is obtained by substituting Equation (2.5) into Equation (2.4) and searching for parameter values that maximize the likelihood of the observed data. There are many excellent treatments of maximum likelihood estimation (Greene, 2003) that can provide the necessary background for researchers in marketing. Maximum likelihood estimation comprises two tasks. The first is a method of finding the maximum using some type of gradient-based hill-climbing technique. The second is an associated method of inference based on large-sample theory for constructing confidence intervals and hypothesis testing.

A second issue associated with Equation (2.5) is what is known as the independence of irrelevant alternative (IIA) property of the logistic function. The IIA property is easy to see by considering the ratio of two choice probabilities:

$$\frac{p_i}{p_k} = \frac{\exp[\beta_{0i} + \beta_p \ln \text{price}_i]}{\exp[\beta_{0k} + \beta_p \ln \text{price}_k]}, \quad (2.6)$$

which does not involve the price or any other aspect of other choice alternatives — e.g., the ratio in (2.6) is not a function of the price of brand j .

Counter-examples of exceptions to IIA abound in the marketing literature, and occur whenever there exists a relatively dissimilar offering in the choice set. Consider, for example, demand from among three well known sodas: 7up, Pepsi, and Coke. As the price of Coke changes, one would expect that the choice probability of Pepsi to be affected more than the choice probability of 7up. Yet, an implication of the IIA property is that choice probabilities are affected in a proportional manner.

So far, we have considered only a very simple choice model based on a multinomial response vector and a logit link function. This model has the following assumptions associated with it:

- (i) Choices are conditionally independent given p .
- (ii) Choice probabilities are driven by parameters that do not change over time — i.e., either p_i is constant, or β_0 and β_p are constant.
- (iii) Demand is represented by zero's and one's, indicating no-choice and choice.
- (iv) There is an explicit set of choice alternatives included in the analysis.
- (v) There is an explicit function form for covariates — i.e., the logarithm of price.
- (vi) Some of the coefficients are unique to the choice alternatives (β_0), while others are constant across choice alternatives (β_p).
- (vii) The IIA property is valid.

These assumptions are all subject to debate, discussion, and extension. They apply in some contexts but not in others. Before progressing with a discussion of approaches to generalizing these assumptions, it is useful to document a selected subset of papers that discuss the existence and analysis of these issues in the marketing literatures. While our brief review leaves out many good papers, it provides a bridge to the choice modeling literature from which an expanded set of published papers can be examined.

2.1 Marketing Literature on Multinomial Logit Models

The logit and the multinomial logit models are the most extensively used choice models in marketing. Logit models were introduced for binary choice by Luce (1959) and its generalization to more than two alternatives led to the multinomial logit model popularized by McFadden (1974). The ease of estimation of logit and multinomial logit models (MNL) was a primary reason for its popularity.

Marketers are often interested in understanding how price, promotions, and other marketing mix variables impact their sales or market share. An initial appeal for the MNL model was due to its being stochastic and yet admitting decision variables like price and promotions (Domencich and McFadden, 1975; Punj and Staelin, 1978; Silk and Urban, 1978; Louviere and Woodworth, 1983). Guadagni and Little (1983) use the MNL model on a scanner panel data to understand the effect of various marketing variables on consumer choice among product alternatives. They demonstrate the statistical significance of the explanatory variables of brand loyalty, size loyalty, store promotions, shelf price, and price cuts on share. Gensch and Recker (1979) compare MNL model to regression models, and demonstrate that MNL is superior for cross-sectional multiattribute choice modeling and understanding household preferences.

There has been much research in marketing addressing the basic and applied assumptions of the MNL models. With regard to the assumptions listed above: (i) McAlister (1982), Kahn (1998), Feinberg et al. (1992) investigate the aspects of variety seeking; (ii) Lattin (1987), Mela et al. (1997), Jedidi et al. (1999), Seetharaman (2004) allow for variation of parameters over time; (iii) purchase quantity is investigated in discrete/continuous models of demand by, Hanemann (1984), Krishnamurthi (1988), Chintagunta (1993), Chiang (1991), Arora et al. (1998); (iv) choice set construction by Erdem and Swait (2004), Gilbride and Allenby (2004); (v) alternative parameterizations by Wagner and Taudes (1986), Vilcassim and Jain (1991), Swait and Louviere (1993), Sonnier et al. (2007); (vi) Krishnamurthi and Raj (1988), Steckel and Vanhonacker (1988), Kamakura and Russell (1989), Bechtel (1990) propose other logit models; and (vii) IIA relaxation by McFadden (1978)

using nested logit, Daganzo (1979), Kamakura and Srivastava (1984, 1986), Chintagunta (1992) using probit and others.

The papers listed above indicate a variety of creative approaches to addressing limitations of the basic multinomial logit model. The goal of our discussion, however, is not to generate a list of published work in the field of choice modeling. The triannual choice symposium initiated by Jordan Louviere, Joel Huber, and others provides a comprehensive link to this literature through reports published in the journal *Marketing Letters*. Instead, we wish to describe a foundational framework for considering models of choice in marketing. We consider future trends once the economic foundation is established.

3

Economic Models of Choice

Economic theory, as with any theory, imposes structure on problems that can be used to evaluate modeling assumptions and guide the development of new models. Models of choice in economics begin with the existence of a scalar measure of consumer utility that can be used to generate a preference ranking of the choice alternatives. Consumers are assumed to make choices that are consistent with the concept of constrained utility maximization:

$$\max u(x) \quad \text{subject to } p'x \leq E, \quad (3.1)$$

where x denotes a vector of quantities, p denotes the vector of prices and E is the budgetary allotment, or total expenditure. E is sometimes called the “income constraint” and is often confused with the notion of a household’s annual income. For logical consistency and a well-defined optimization problem, E can only refer to the budget allocated to the collection of alternatives under study. The implicit conditioning introduced by the decision to study choices in a particular category only cannot be ignored in the formulation of the constraint. Thus, a better interpretation for E is the maximum expenditure a consumer is willing to make in the product category, which is different from the annual

household income. This implies the observed choices of a consumer are affected by the budgetary allotment regardless of the magnitude of the prices, affecting both low priced items such as toothbrushes and high priced items like automobiles.

Economic utility functions have the following properties — they are positively valued, (weakly) increasing in x , and have nonincreasing marginal returns (e.g., diminishing marginal returns). There are instances in the marketing literature where utility functions are said to have a “ U ” shape, where consumers prefer either small quantities or large quantities, and instances where utility functions are assumed to have an inverted “ U ” shape where moderate amounts of quantity are preferred. These instances refer to what is known as the indirect utility function, not the utility function. The indirect utility function is obtained by solving for the maximum attainable utility given the budgetary allotment. Below we show that most choice models in marketing lack budgetary allotment implications, making it easy to mischaracterize utility functions as having either “ U ” or an inverted “ U ” shape.

To illustrate, consider a constrained maximization problem involving the Stone–Geary (Parker, 2000) utility function:

$$\begin{aligned} \max u(x) = & \beta_1 \ln(x_1 + \gamma_1) \\ & + \beta_2 \ln(x_2 + \gamma_2) \quad \text{subject to } p_1 x_1 + p_2 x_2 \leq E \end{aligned} \quad (3.2)$$

in which the β 's and γ 's are parameters. The β parameters are assumed to sum to one (i.e., $\beta_1 + \beta_2 = 1.0$) for identification purposes. Restricting the β parameters ensures that different β and γ parameter values are associated with different preference orderings, ruling out the scaling of utility by an arbitrary factor (e.g., $\kappa u(x)$ and $u(x)$ implying the same ranking of goods). We can solve for the utility maximizing quantities x^* by first forming an auxiliary function (see Simon and Blume, 1994):

$$L = \beta_1 \ln(x_1 + \gamma_1) + \beta_2 \ln(x_2 + \gamma_2) - \lambda(p_1 x_1 + p_2 x_2 - E), \quad (3.3)$$

where λ is known as the Lagrangian multiplier. Differentiating with respect to x , we obtain a set of first-order conditions:

$$\begin{aligned} 0 &= \frac{\partial L}{\partial x_1} = \frac{\beta_1}{(x_1 + \gamma_1)} - \lambda p_1 \Leftrightarrow \beta_1 = \lambda p_1 (x_1 + \gamma_1), \\ 0 &= \frac{\partial L}{\partial x_2} = \frac{\beta_2}{(x_2 + \gamma_2)} - \lambda p_2 \Leftrightarrow \beta_2 = \lambda p_2 (x_2 + \gamma_2). \end{aligned} \quad (3.4)$$

Using the constraint $\beta_1 + \beta_2 = 1.0$, we can solve for λ :

$$\lambda = \frac{1}{E + p_1 \gamma_1 + p_2 \gamma_2}, \quad (3.5)$$

and substituting back into the first-order conditions allows us to solve for the optimal demand (x^*) equations:

$$\begin{aligned} x_1^* &= -\gamma_1 + \frac{\beta_1}{p_1} (E + p_1 \gamma_1 + p_2 \gamma_2), \\ x_2^* &= -\gamma_2 + \frac{\beta_2}{p_2} (E + p_1 \gamma_1 + p_2 \gamma_2). \end{aligned} \quad (3.6)$$

Finally, the expressions for optimal demand can be substituted into the utility function to obtain the expression for indirect utility — i.e., the maximum attainable utility as a function of prices (p) and expenditure (E):

$$u(x^*) = \sum_{j=1}^2 \beta_j (\ln(\beta_j) - \ln(p_j) + \ln(E + p_1 \gamma_1 + p_2 \gamma_2)). \quad (3.7)$$

It is important to notice that the direct utility function in Equation (3.2) does not contain any price terms, whereas the indirect utility function in Equation (3.7) does. Researchers in marketing often incorrectly refer to functions with price terms as “utility function.” In reality, these correspond to indirect utility functions that are neither increasing nor possess diminishing marginal returns.

In our discussion below, we discuss economic choice models in terms of direct utility, not indirect utility. There are two reasons for this. First, we believe it is more natural to build models in terms of direct utility that describes preferences and their formation. Moreover, models based on a direct utility specification are easier to estimate with modern

Bayesian methods, even when closed-form expressions of demand and indirect utility are not available. Thus, generalizations of models such as the Stone–Geary, can be more readily explored. We begin our discussion of economic choice models by describing the economic assumptions that would lead to the multinomial-logit model of Equation (2.5). We then explore departures from these assumptions and their implications for choice modeling.

3.1 The Economics of Discrete Choice

Discrete choice refers to outcomes that represent purchases of discrete goods, such as packaged goods in a supermarket. When making a selection from among branded alternatives in a product category, consumers often buy just one alternative — e.g., one brand of peanut butter, one brand of hot dog, or one brand of margarine. In all these cases, the choice alternatives are near-perfect substitutes and the solution to Equation (3.1) involves just one element of x being nonzero. A utility function that gives rise to discrete choice has marginal utilities that are constant:

$$u(x) = \sum_{k=1}^K \psi_k x_k = \psi'x, \quad (3.8)$$

where ψ_k is the marginal utility for alternative k — i.e., $\partial u(x)/\partial x_k$. The utility function in (3.8) represents a boundary condition for utility functions — it is positive as long as the elements of ψ are positive, increasing, and has a nonincreasing first derivative.

Figure 3.1 displays a contour plot of Equation (3.8) for the case where $K = 2$, $\psi_1 = 1$, and $\psi_2 = 3$. The solid lines on the plot represent lines of constant utility, referred to as indifference curves in the economics literature. The indifference curves are linear and parallel, with slopes that do not change as they get further from the origin. The dotted line in Figure 3.1 represents the budget constraint $p_1 x_1 + p_2 x_2 = E$, with $p_1 = 1$, $p_2 = 2$, and $E = 6$. Since both the indifference curves and budget constraint are linear, utility is maximized by selecting just one of the alternatives, not both. In Figure 3.1, utility is maximized at the point $x' = (0, 3)$.

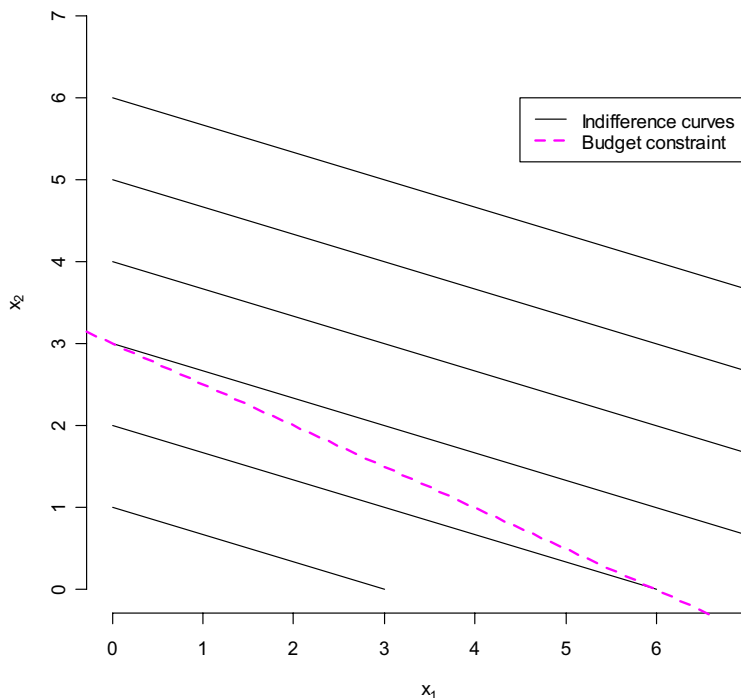


Fig. 3.1 Linear utility contours and budget constraint. Contours (indifference curves) further from the origin are associated with higher levels of utility. Changes in expenditure results in translations of the budget constraint, but not in the good associated with utility maximization.

The general solution to constrained utility maximization is to select the choice alternative for which the ratio ψ_j/p_j is maximum. For the example in Figure 3.1, the ratio is equal to 1.0 for the first choice alternative, and equal to 1.5 for the second choice alternative. Thus, the second choice alternative provides a greater bang-for-the-buck, and the consumer is better off selecting just the second item than any combination of the first and second item. This is known as a “corner solution” in the economics literature, as opposed to an “interior solution” where utility is maximized with nonzero quantities for each element of the vector x .

There are a number of interesting aspects to the utility maximizing solution using the utility function in Equation (3.8). First, the solution is expressed as the brand identified with largest bang-for-the-buck

(ψ_j/p_j) and it does not involve the budgetary allotment E . Thus, the solution identifies which alternative to purchase, and the quantity purchased is computed as $x_j = E/p_j$. As expenditure, E , changes, the utility maximizing solution changes only in terms of the optimal quantity, x_j , and not the brand leading to utility maximization. This assumption of the model is violated in many contexts, offering opportunities for alternative choice models as discussed below.

Second, the solution is a special case of what is known in the economics literature as Kuhn–Tucker conditions. These are general conditions that are used to associate observed demand to model parameters under the assumption of constrained utility maximization. The conditions are derived using basic principles of constrained maximization by first forming an auxiliary function:

$$L = u(x) - \lambda(p'x - E). \quad (3.9)$$

Differentiating the auxiliary function leads to the Kuhn–Tucker first-order conditions:

$$\begin{aligned} u_j - \lambda p_j &= 0 & \text{if } x_j^* > 0 \\ u_j - \lambda p_j &< 0 & \text{if } x_j^* = 0, \end{aligned} \quad (3.10)$$

where x^* is the vector of observed optimal demand, and u_j is the derivative of the utility function with respect to x_j . Applying the Kuhn–Tucker conditions in (3.10) to the constant marginal utility model in (3.8) gives us conditions that link observed demand to the underlying economic model.

$$\text{if } x_j^* > 0 \quad \text{then } \frac{\psi_j}{p_j} > \frac{\psi_k}{p_k} \quad \text{for all } k. \quad (3.11)$$

Note that this system does not have an interior solution. The case where $u_j - \lambda p_j = u_i - \lambda p_i = 0$ is degenerate in that any combination x_i and x_j meeting the budget constraint is possible. This case will be ruled out next.

Equation (3.11) serves as the basis for formulating an economic model of discrete choice. An important aspect of these models is the introduction of error terms that allow the observed data to depart probabilistically from the deterministic choices implied by (3.11). A standard assumption is that consumers act according to Equation (3.11),

but that on any given choice occasion a person's marginal utility may vary:

$$\psi_{j,t} = \psi_j e^{\varepsilon_{j,t}}, \quad (3.12)$$

where $\varepsilon_{j,t}$ is viewed by the analyst as a random element that leads to variation in the consumer's marginal utility over time or across choices, where $E[\varepsilon_{j,t}] = 0$ and σ_j is the scale parameter for $\varepsilon_{j,t}$. The error term is exponentiated to ensure that it takes on only positive values so that Equation (3.8) remains a valid utility function. The introduction of an error term implies that the analyst cannot perfectly predict what a consumer will choose. Instead, choice probabilities are formed based on Equations (3.11) and (3.12):

$$\begin{aligned} p_i &= \Pr(x_i^* > 0) \\ &= \Pr\left(\frac{\psi_{i,t}}{p_{i,t}} > \frac{\psi_{k,t}}{p_{k,t}} \text{ for any } k \neq i\right) \\ &= \Pr(\ln \psi_i - \ln p_{i,t} + \varepsilon_{i,t} > \ln \psi_k - \ln p_{k,t} + \varepsilon_{k,t} \text{ for any } k \neq i). \end{aligned} \quad (3.13)$$

Thus, choice probabilities are obtained by integrating over regions of the ε space. For K choice alternatives, the errors $\{\varepsilon_{j,t}\}$ span a K dimensional space, and regions of that space map onto the condition that $\psi_{j,t}/p_j$ is maximum. If we assume that the error terms are distributed identically (e.g., $\sigma_j = \sigma$) and independently (e.g., $\pi(\varepsilon_i, \varepsilon_j) = \pi(\varepsilon_i)\pi(\varepsilon_j)$), then the choice probabilities can be expressed as (suppressing the subscript t):

$$\begin{aligned} p_i &= \Pr(x_i^* > 0) \\ &= \Pr(\ln \psi_i - \ln p_i + \varepsilon_i > \ln \psi_k - \ln p_k + \varepsilon_k \text{ for any } k \neq i) \\ &= \Pr(V_i + \varepsilon_i > V_k + \varepsilon_k \text{ for any } k \neq i) \\ &= \Pr(\varepsilon_k < V_i - V_k + \varepsilon_i \text{ for any } k \neq i) \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{V_i - V_1 + \varepsilon_i} \cdots \int_{-\infty}^{V_i - V_k + \varepsilon_i} \pi(\varepsilon_k) \cdots \pi(\varepsilon_1) \right] \pi(\varepsilon_i) d\varepsilon_k \cdots d\varepsilon_1 d\varepsilon_i \\ &= \int F(V_i - V_1 + \varepsilon_i) \cdots F(V_i - V_k + \varepsilon_i) \pi(\varepsilon_i) d\varepsilon_i, \end{aligned} \quad (3.14)$$

where $F(\cdot)$ denotes the cumulative density function (cdf) of ε , and $\pi(\cdot)$ denotes the probability density function (pdf). This is a formidable calculation. However, if the error term ε is distributed extreme value type I, we have $F(\varepsilon) = \exp[-e^{-\varepsilon/\sigma}]$ and this calculation has a closed-form expression (McFadden, 1974):

$$\begin{aligned} p_i = \Pr(x_i^* > 0) &= \frac{\exp\left[\frac{\ln \psi_i - \ln p_{i,t}}{\sigma}\right]}{\sum_{j=1}^K \exp\left[\frac{\ln \psi_j - \ln p_{j,t}}{\sigma}\right]} \\ &= \frac{\exp[\beta_{0,i} - \beta_p \ln p_{i,t}]}{\sum_{j=1}^K \exp[\beta_{0,j} - \beta_p \ln p_{j,t}]}, \end{aligned} \quad (3.15)$$

where $\beta_{0,i} = \ln \psi_i / \sigma$ and $\beta_p = 1/\sigma$. In other words, the price coefficient is the inverse of the scale of the error term. As the random component of utility becomes bigger, the sensitivity of the choice probability to price changes becomes smaller. In addition, the intercepts β_0 are a function of both log marginal utility and the scale of the error term (Swait and Louviere, 1993).

It is important to remember, though, that the expression in the exponent of Equation (3.15) is not the utility of a specific choice alternative. These expressions are related to the utility function through the ψ terms, but also include price and the scale value of the error term. Equation (3.15) is better thought of as the demand function, similar to Equation (3.6).

The above discussion demonstrates how the multinomial logit model in Equation (2.5) can be considered as an economic choice model with constant marginal utility and extreme value errors. The advantage of understanding the relationship between Equations (2.5) and (3.15) is that it answers the specific question of “why use the logarithm of prices?” In doing so, however, it also raises a number of questions that serve as the basis for additional research:

- (i) What about other error term distributions?
- (ii) What about quantity?
- (iii) Why constant marginal utility?
- (iv) What if per-unit prices decline with quantity purchased?

3.2 Other Error Distributions and State Dependence

The extreme value error term leads to a closed form expression for the choice probability that is identical to one that directly assumes a multinomial response and a logistic function for parameterizing the choice probabilities. Other distributions lead to other forms for the choice probabilities. For example, a generalization of the extreme value distribution (GEV) introduces block-correlation in a manner similar to that found in an intra-class structure (Anderson, 1971; McFadden, 1978). The GEV has a distribution of the form:

$$F(\varepsilon_1, \dots, \varepsilon_k) = \exp \left\{ - \sum_{s=1}^S \left[\sum_{j \in A_s} \exp[-\varepsilon_j / \lambda_s] \right]^{\lambda_s} \right\}, \quad (3.16)$$

where $A = A_1 \cup A_2 \cup \dots \cup A_S = \{1, 2, \dots, k\}$ represents the partitioning of the k offerings into S partitions A_1, A_2, \dots, A_S , and λ_s is a measure of the relative independence of the alternatives within each set. If $\lambda_s = \lambda = 1$, then the cumulative density function in (3.16) is equal to the product of standard extreme value cumulative density functions. Since the arguments within each partition are exchangeable, the distribution has an equi-correlated structure within group, and an equi-correlated structure across groups.

McFadden (1984) showed that the use of GEV distribution leads to choice probabilities of the form:

$$\begin{aligned} p_i &= \Pr(x_i^* > 0) = \Pr(\varepsilon_j \leq V_i - V_j + \varepsilon_i, j \in A) \\ &= \int_{-\infty}^{+\infty} F_i(V_i - V_1 + \varepsilon_i, \dots, V_i - V_k + \varepsilon_i) d\varepsilon_i \\ &= \frac{\exp[V_i / \lambda_{s'}]}{\sum_{j \in A_{s'}} \exp[V_j / \lambda_{s'}]} \cdot \frac{\left[\sum_{j \in A_{s'}} \exp[V_j / \lambda_{s'}] \right]^{\lambda_{s'}}}{\sum_{s=1}^S \left[\sum_{j \in A_s} \exp[V_j / \lambda_s] \right]^{\lambda_s}}, \end{aligned} \quad (3.17)$$

where alternative i is an element of partition $A_{s'} (i \in A_{s'})$, and F_i denotes the derivative of the joint cdf with respect to element i . The second line in Equation (3.17) is a generalization of the last

line of Equation (3.14) for the dependent case — i.e., $F_3(x_1, x_2, x_3) = F(x_1)F(x_2)\pi(x_3)$ for independently distributed random variables. The choice probability in (3.17) has a nested structure equal to the probability of selecting the alternative within the partition, multiplied by the probability of selecting the partition: $p_i = \Pr(i|A_{s'}) \cdot \Pr(A_{s'})$.

Researchers in marketing have employed the so called nested logit model in various ways, often motivating the model as a series of sequential decisions. However, it is important to realize that the model is equally motivated for consumers engaged in utility maximization with correlated extreme value error terms (Kannan and Wright, 1991; Allenby and Rossi, 1993).

While the extreme value distribution plays a central role in statistical theory as the distribution of the minimum and maximum of a set of random variables (Johnson and Kotz, 1971), a more commonly observed distribution in the marketing literature is the normal distribution, which, by the Central Limit Theorem, is the distribution of the mean of a set of random variables. The use of the normal distribution leads to a probit model for discrete choice. An advantage of using the normal distribution is it admits a general correlational structure and thus flexibly relaxes the IIA property of the logit model. A disadvantage of the normal distribution is that, prior to the advent of modern Bayesian computing, the computation of the integral in Equation (3.14) was challenging because the choice probabilities did not have a closed form.

The Bayesian method of data augmentation (Tanner and Wong, 1987, see also Rossi et al., 2005) has revolutionized the analysis of latent models by avoiding the direct evaluation of the integral in Equation (3.14). In choice models, the method involves the generation of latent random utilities. Bayes theorem is used to derive the conditional distribution of the utilities, and Monte Carlo draws are obtained and used in a Markov Chain with equilibrium distribution equal to the posterior distribution of model parameters.

Apart from the complexity of computing high-dimensional integrals, researchers have also found parameter identification to be extremely tenuous in a Multinomial Probit (MNP) models (Keane, 1992). As we will see below, Bayesian methods have opened the

door for estimation and exploration of a wide variety of choice models, including the MNP model (Rossi et al., 2005). The use of the Bayesian method of data augmentation has led to development of various complex behavioral and economic models that help better understand consumer decision processes. For example: (i) Gilbride and Allenby (2004) investigate consumers' use of screening rules as part of a discrete choice model using a standard hierarchical Bayes probit model and find that more than 90% of the respondents screen alternatives on one or more attributes. (ii) Wedel and Pieters (2000) propose a probit type model to understand consumer eye fixations on print advertisements that lead to memory for advertised brands. They adopt a Bayesian approach to estimate and investigate the model's assumptions. (iii) Hruschka (2007) uses an MNP model that combines heterogeneity across households with a highly flexible deterministic utility function (approximated by a neural net of the multilayer perceptron type). He demonstrates that the MNP model extended by a multilayer perceptron outperforms all other models investigated.

In addition, researchers in marketing have captured state dependence by (i) allowing alternative specific utility errors to be serially correlated over time (Allenby and Lenk, 1994, 1995), (ii) modeling carryover effects, habit persistence, inertia or variety seeking behavior between adjacent choice observations (Erdem, 1996; Roy et al., 1996; Keane, 1997), and (iii) incorporating distributed lag specifications of the preference parameters that are impacted by various marketing actions (Jedidi et al., 1999; Seetharaman, 2004). Lachaab et al. (2006) propose a general state space heterogeneous MNP model where they demonstrate that accounting for both consumer heterogeneity and temporal dependence outperforms most of the models that ignore either of these aspects.

3.3 Outside Goods, Budget Constraints, and the No-Choice Option

Multinomial choice models implicitly assume that one and only one of the choice alternatives is selected on a purchase occasion. Consumers,

however, frequently decide to purchase none of the alternatives due to factors such as sufficient inventory at home, or a decision to postpone purchase and wait for a better deal. A simple way of incorporating no-choice into the multinomial model is to simply associate no-choice with one of the multinomial outcomes. The problem with this, however, is that it is difficult to characterize the no-choice option — i.e., what value should be given to $\psi_{\text{outside}} = \exp[V_{\text{outside}}]$?

Traditionally, researchers in marketing have assumed that $V_{\text{outside}} = 0.0$, or $\psi_{\text{outside}} = 1.0$, and have assigned a price of 1.0. The reason for these assignments can be seen in Equation (3.11). Choice is driven by the ratio of marginal utility to price, and these assignments serve to identify the remaining marginal utilities (ψ) and scale of the error term ($1/\sigma = \beta_{\text{price}}$). It is the ratio of marginal utilities to prices among the choice alternatives that determines choice, not their absolute values. This can be seen by re-writing Equation (3.11) as

$$\text{if } x_j^* > 0 \quad \text{then } \frac{\psi_j}{\psi_k} > \frac{p_j}{p_k} \quad \text{for all } k. \quad (3.18)$$

Thus, some form of standardization is needed in all choice models, even those with a no-choice option, and the practice of setting $\psi_{\text{outside}} = 1.0$ does not impair the analysis.

A more formal treatment of the no-choice option is to include an outside good in the constrained maximization problem in Equation (3.1):

$$\max u(x, z) \quad \text{subject to } p'x + z \leq E, \quad (3.19)$$

where z denotes the outside good with a price of 1.0. Consumers are now assumed to divide their budgetary allotment among choice alternatives within the product category and the outside good, and “no-choice” is equivalent to allocating all of their expenditure to the outside good.

A simple utility function that guarantees a discrete choice outcome is the linear utility structure in Equation (3.8):

$$u(x, z) = \sum_{j=1}^K \psi_j x_j + \psi_z z = \psi'x + \psi_z z, \quad (3.20)$$

which leads immediately to the multinomial choice model with solution described by Equation (3.15). With linear utility, the budgetary

allotment does not enter the utility maximizing solution and the model does not formally deal with the quantity purchased.

Some researchers consider applying a general utility function to a domain where demand for the inside good is restricted to take on values of zero or one with just one alternative chosen, and the remaining allotment allocated to the outside good (Jedidi and Zhang, 2002; Jedidi et al., 2003). These models do not rely on the Kuhn–Tucker conditions (Equation (3.4)) for model estimation, and instead enumerate all possible solutions to the utility maximization knowing that there are just $K + 1$ possibilities. The utility associated with choosing alternative j is then:

$$u(x_j = 1, z_j) = \psi_j + \alpha_z(E - p_j), \quad (3.21)$$

where α_z is the marginal utility of the outside good. In this formulation the term $\alpha_z E$ is common to all outcomes and does not enter the utility maximizing solution.

For the budgetary allotment to play a meaningful role in the solution to the utility maximizing problem, it is necessary to introduce some form of nonlinearity into the utility function. One way of doing this is to assume a multiplicative relationship between alternatives in the product category and the outside good, as in a Cobb–Douglas utility function:

$$\begin{aligned} \ln u(x, z) &= \alpha_0 + \alpha_x \ln u(x) + \alpha_z \ln(z) \\ u(x) &= \psi'x. \end{aligned} \quad (3.22)$$

There are two utility functions present in Equation (3.22). The first equation is a bivariate utility function between goods in the product class and an outside good. The second equation describes a sub-utility among items within the product class. Because the sub-utility function is linear, just one of the choice alternatives within the product category (x) will have nonzero demand. Maximizing the utility function in Equation (3.22) subject to the budget constraint $p'x + z \leq E$ is accomplished by substituting $E - p'x$ for z and searching for the (x, z) combination that maximizes utility. This search does not make use of the Kuhn–Tucker conditions, but is a feasible solution strategy because of the presence of the linear sub-utility function. Linear sub-utility

implies that the alternatives within the product class are near-perfect substitutes, and just one of the alternatives will be selected. Thus, the search over all possible (x, z) combinations is simplified to the set of outcomes $\{(x_j, z), j = 1, \dots, K\}$. That is, interior solutions in the x space do not occur.

The solution strategy when only one element of x is nonzero involves two steps. The first step involves identifying the quantity of each brand that maximizes the utility function. The second step involves a comparison among these maximum values for the brands under study. In conducting these two steps, it is important to recognize the presence of error terms in the model. Consistent with our earlier discussion, we assume that marginal utility, ψ_j , contains a multiplicative error term as in Equation (3.12): $\psi_{j,t} = \psi_j e^{\varepsilon_{j,t}}$. Thus, as shown below, the first step becomes a deterministic search among different quantities, while the second step involves a stochastic evaluation of the maximum similar to the standard discrete choice model.

The first step involves finding the quantity that maximizes $u(x_j, z)$ for each brand j :

$$\begin{aligned} \hat{x}_j &= \arg \max_{x_j} \{ \alpha_0 + \alpha_x \ln(\psi_{j,t} x_j) + \alpha_z \ln(E - p_j x_j) \} \\ &= \arg \max_{x_j} \{ \alpha_0 + \alpha_x \ln(\psi_j) + \alpha_x \ln(x_j) + \alpha_x \varepsilon_{j,t} + \alpha_z \ln(E - p_j x_j) \} \\ &= \arg \max_{x_j} \{ \alpha_x \ln(x_j) + \alpha_z \ln(E - p_j x_j) \}, \end{aligned} \quad (3.23)$$

where the simplification in the third line of the equation occurs because terms common across all values of x_j do not affect the result. The second step involves a comparison of utilities evaluated at the \hat{x}_j values. In this second step the error does not cancel, and for extreme value errors we have:

$$\Pr(x_j^*) = \frac{\exp[\alpha_0 + \alpha_x \ln \psi_j + \alpha_x \ln \hat{x}_j^* + \alpha_z \ln(E - p_j \hat{x}_j^*)]}{\sum_{k=1}^K \exp[\alpha_0 + \alpha_x \ln \psi_k + \alpha_x \ln \hat{x}_k + \alpha_z \ln(E - p_k \hat{x}_k)]}. \quad (3.24)$$

The notation “ \wedge ” appearing in the denominator of Equation (3.24) for the brands not selected — they pertain to the unobserved utility maximizing quantities conditional on the estimated parameters.

Choice models in which the budgetary allotment (E) plays a meaningful role typically require the analysis of the purchase quantity (x). An alternative formulation is to assume the existence of a reservation value for which nonzero demand occurs. If the value for all choice options, defined as ψ_j/p_j are below the reservation value, then no-choice occurs, and if at least one value is above the threshold then nonzero demand is optimal. Thus, the choice equation becomes:

$$p_i = \Pr(x_i^* > 0) = \Pr\left(\frac{\psi_{i,t}}{p_{i,t}} > \left(\frac{\psi_{k,t}}{p_{k,t}}, \alpha_{RP}\right) \text{ for any } k \neq i\right), \quad (3.25)$$

where α_{RP} is a parameter for the reservation price. No choice occurs when all the ratios $\psi_j/p_j < \alpha_{RP}$.

Chiang (1991) uses Equation (3.25) to study the nonpurchase decision of individuals, deriving the nonpurchase and purchase probabilities of a given brand using extreme value errors. Chib et al. (2004) proposed a model in which the no-purchase decision depends on price, feature, and display of each brand in the category and on the household inventory. Their model associates no-purchase through marketing-mix covariates and through unobservables that affect both outcomes. They demonstrate their model to be: (i) more distinct from a restrictive specification in which the no-purchase outcome is modeled as an additional outcome (Chiang, 1991; Chintagunta, 2002); (ii) more general than most of the other translog utility models (Chintagunta, 1993; Arora et al., 1998) by allowing unobserved correlations between category-purchase and brand choice decisions.

3.4 Superior and Normal Goods

The constant marginal utility function associated with standard choice models leads to the outcome that just one of the product offerings is demanded at any time. The constant marginal utility function also implies that, as the budgetary allotment, E , increases, the utility maximizing solution does not change. The reason for this property can be seen by inspecting Equation (3.10). The Kuhn–Tucker conditions associated with optimal choice involve marginal utilities and prices. If expenditure shares (i.e., $p_j x_j / E$) do not change as the level of expenditure or maximum attainable utility increases, then the utility

maximizing solution is homothetic — i.e., utility maximizing solutions represent a ray from the origin of the demand space where increases in E simply lead to purchasing greater quantities in the same proportion.

The homothetic property of the utility function is especially problematic for discrete choice models. As the budgetary allotment increases, the utility maximizing solution indicates that consumers will simply purchase more of the same offering they preferred at lower levels of expenditure. In reality, consumer will trade up to higher quality offerings. This is true in product categories ranging from ice cream to luxury automobiles (Allenby et al., 2008).

In many instances, it is desirable to retain aspects of a discrete choice model while allowing for the possibility of superior goods. For example, consumers may purchase only one of a variety of offerings in categories such as bicycles, vacations and electric razors, and choice is still characterized as a strict corner solution where just one of the alternatives has nonzero demand. As demonstrated earlier, this can only be ensured when indifference curves are linear.

Allenby and Rossi (1991) and Allenby et al. (2008) propose an implicitly defined utility function with linear indifference curves but nonconstant marginal utility:

$$\begin{aligned} \ln u(x, z) &= \alpha_0 + \alpha_x \ln u(x) + \alpha_z \ln(z) \\ u(x) &= \sum_{k=1}^K \psi_k(\bar{u}) x_k = \sum_{k=1}^K \exp[\beta_k - \kappa_k \bar{u}(x, z)] x_k, \end{aligned} \quad (3.26)$$

where marginal utility, ψ , is a function of attainable utility \bar{u} . Attainable utility increases and the relative values of ψ change as respondents allocate greater expenditure to the product class. The utility function is linear in x , indicating that the indifference curves are linear. Moreover, for $\kappa > 0$, it can be shown that the indifference curves do not intersect in the positive orthant, and Equation (3.26) is a valid utility function — i.e., is a positive nondecreasing function with diminishing marginal returns. The ratio of marginal utilities, e.g., ψ_i/ψ_j , is a function of \bar{u} such that values of κ closer to zero are associated with superior goods — they are more preferred as \bar{u} is larger, which coincides with larger expenditure, E .

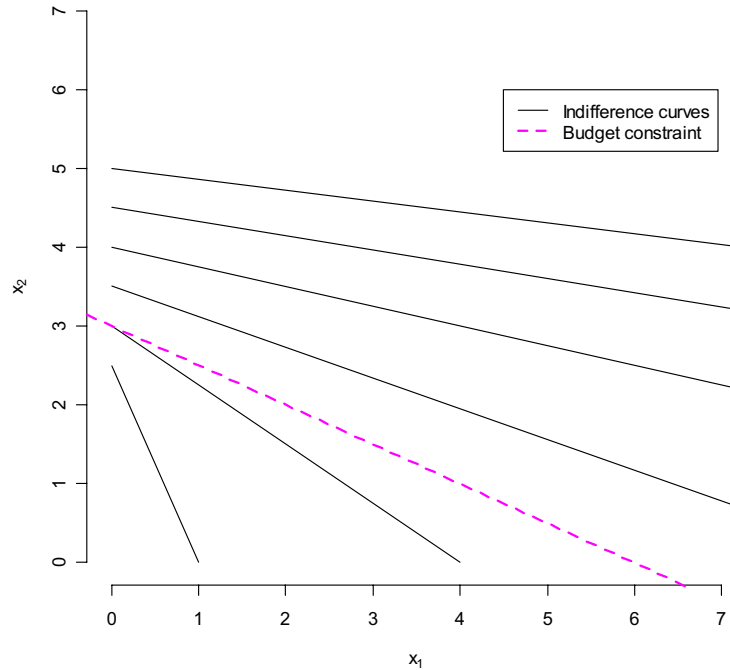


Fig. 3.2 Non-homothetic linear utility contours and budget constraint. Contours (indifference curves) further from the origin are associated with higher levels of utility. Higher levels of expenditure are associated with preference for good 2 (vertical axis).

An illustration of choice for the utility function in (3.26) is provided in Figure 3.2. The solid lines in the figure correspond to indifference curves that fan out with good x_2 superior and x_1 relatively inferior. At low levels of expenditure, E , the utility maximizing solution corresponds to the selection of positive quantities of x_1 . At higher levels of expenditure, x_2 is preferred. It should be remembered that an “income effect” can be produced either through a relaxation of the expenditure, or through changes in the relative prices of the offerings that would induce a change in the slope of the budget constraint. As consumers increase their attainable level of utility, good two is preferred to good one.

The evaluation of choice probabilities for estimation of model parameters proceeds in a manner similar to the direct search algorithm described above. Since the sub-utility function for goods in the product

class is associated with linear indifference curves, there will be just one element of x with nonzero demand, a strict corner solution. Thus, it is feasible to engage in a direct search procedure where utility is evaluated at each solution point along the boundary of the budgetary allotment, i.e., $p'x + z = E$. If we restrict the utility function to the evaluation of utility with quantities of either zero or one, the attainable utility at each corner is obtained by solving the implicit function:

$$\begin{aligned} \ln u^j &= \beta_j - \kappa_j u^j + \alpha_z \ln(E - p_j) \\ u^j &= u(x_j = 1, z = E - p_j). \end{aligned} \tag{3.27}$$

We omit α_0 and α_x from (3.26) because the data are only informative about utility ratios and quantities are restricted to be either zero or one. Choice probabilities can then be obtained by introducing an error term in the expression for marginal utility ($\beta_{jt} = \beta_j e^{\varepsilon_{jt}}$) and determining which of the choice alternatives result in maximum utility (see Equation (3.24)).

It is useful to compare Equation (3.27) with (3.21). In Equation (3.21), expenditure (E) enters linearly into the indirect utility function and plays no role in choice probabilities because it contributes equally to all choices. In Equation (3.27), expenditure enters nonlinearly, primarily because utility is implicitly defined (i.e., u^j appears on both the left and right side of the equation). The solution to the implicit function is obtained using numerical methods, such as Newton's method.

Other models with asymmetric effects rely on heterogeneity and noneconomic models to reflect the tendency of individuals to trade-up to higher quality goods. Blattberg and Wisniewski (1989), for example, explain trade-up behavior through the distribution of preferences for quality. Wedel et al. (1995) employ piece-wise exponential hazard models, coupled with heterogeneity, to account for the presence of asymmetry in brand switching.

3.5 Satiation

The general solution to the consumer task of maximizing utility subject to a budget constraint involves interior solutions with optimal demand

taking on values different from just zero and one. In the analysis of packaged goods, demand for nonnegative integer values is possible, and in service categories (e.g., cell phone usage) demand can be a continuous variable.

Much of the applied demand literature is characterized by interior solutions where demand is greater than zero for all items. The reason is because this literature operates at a high level of aggregation, either by studying demand for all brands within a product category or by applying economic models to aggregation of individual consumers. The marketing literature, however, is more focused on the behavior of individual respondents in specific contexts, and must deal directly with the possibility of corner solutions where at least one of the offerings has zero demand.

An example of a utility function capable of dealing with corner and interior solutions was proposed by Kim et al. (2002):

$$u(x) = \sum_{k=1}^K \psi_k (x_k + \gamma_k)^{\alpha_k}, \quad (3.28)$$

where k is an index of the choice alternatives, x is the vector of demand, ψ_k is a baseline utility parameter, and α_k is parameter that ensures diminishing marginal returns when it is restricted to the unit interval.

The term γ_k is an offset parameter that translates the utility function so that interior and corner solutions are possible. For $\gamma > 0$, the indifference curves cross the axes and allow for corner solutions. Figure 3.3 displays the indifference curves when $\psi' = (1.0, 2.0)$, $\alpha' = (0.6, 0.3)$, and $\gamma' = (1.0, 1.0)$.

The estimation of the parameters in Equation (3.28) requires the use of the Kuhn–Tucker conditions to relate observed demand x^* to the assumed process of utility maximization. In Figure 3.3, the x^* is equal to the point at which the budget constraint just touches the indifference curves. The presence of interior solutions makes it impractical to search over all possible solution points as was possible with a discrete choice model, or models where demand was known to reside along one of the axes. When interior solutions are present, the likelihood of observed demand is a function of multiple error terms as explained below. Corner solutions correspond to inequality restrictions that give rise to mass

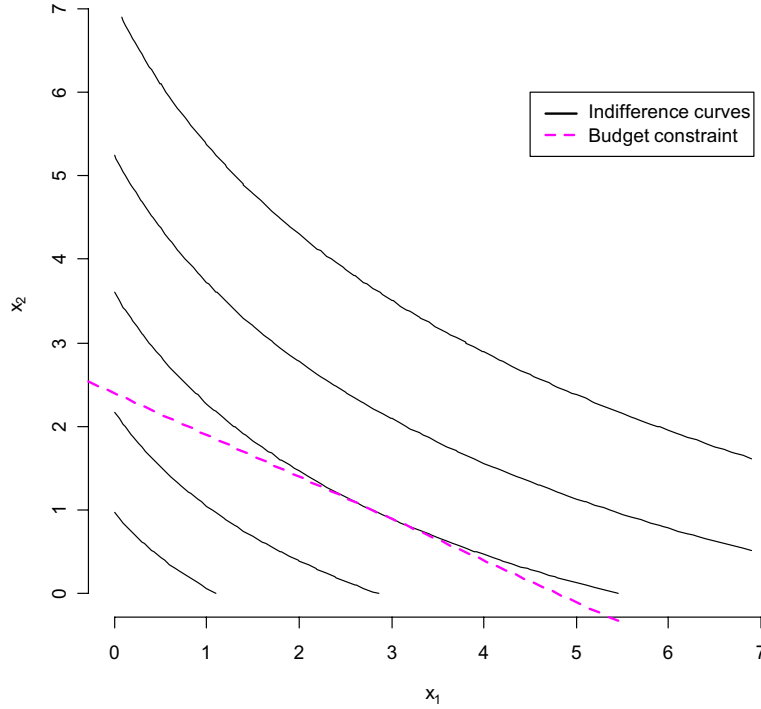


Fig. 3.3 Utility contours with diminishing marginal returns. Contours (indifference curves) further from the origin are associated with higher levels of utility. Price changes can induce changes in the budget constraint to produce either interior (shown) or corner solutions because indifference curves do not asymptote at zero quantities.

points of probability. Interior solutions correspond to a condition where marginal utility divided by marginal price, or the “bang-for-the-buck,” is equal among the options with nonzero demand. This gives rise to a density contribution to the likelihood function.

Marginal utility can be calculated from Equation (3.22) as $u_i = \alpha_i \psi_i (x_i + \gamma_i)^{\alpha_i - 1}$, and the Kuhn–Tucker conditions imply that

$$\frac{u_i}{p_i} = \frac{u_j}{p_j} \quad \text{for } x_i^* > 0, x_j^* > 0 \quad (3.29)$$

and

$$\frac{u_i}{p_i} > \frac{u_k}{p_k} \quad \text{for } x_i^* > 0, x_k^* = 0. \quad (3.30)$$

Details of estimating demand models with Equations (3.29) and (3.30) are provided by Kim et al. (2002), and are repeated here for completeness (see also Pudney (1989, Chap. 4)). We derive the likelihood by introducing error into the baseline parameter as in Equation (3.12). Taking logs of the Kuhn–Tucker conditions, we have:

$$V_j(x_j^*|p) + \varepsilon_j = \ln \lambda \quad \text{if } x_j^* > 0 \quad (3.31)$$

$$V_j(x_j^*|p) + \varepsilon_j < \ln \lambda \quad \text{if } x_j^* = 0, \quad (3.32)$$

where λ is Lagrange multiplier and $V_j(x_j^*|p) = \ln(\psi_j \alpha_j (x_j^* + \gamma_j)^{\alpha_j - 1}) - \ln(p_j)$ $j = 1, \dots, m$.

Optimal demand satisfies the Kuhn–Tucker conditions in (3.31) and (3.32) as well as the “adding-up” constraint induced by the budget, $p'x^* = E$, which induces a singularity in the distribution on x^* . This singularity was dealt with in Equation (3.21), for example, by substituting $z = E - p'x$. Since, in our specification, at least one alternative will always be chosen, we assume that the first good is always purchased (i.e., this involves a simple re-labeling of the alternatives) and subtract Equation (3.31) from the others. This reduces the dimensionality of the system of equations by one. Equations (3.31) and (3.32) are now equivalent to:

$$\nu_j = h_j(x^*, p) \quad \text{if } x_j^* > 0 \quad (3.33)$$

$$\nu_j < h_j(x^*, p) \quad \text{if } x_j^* = 0, \quad (3.34)$$

where $\nu_j = \varepsilon_j - \varepsilon_1$ and $h_j(x^*, p) = V_1 - V_j$ and $j = 2, \dots, m$.

The likelihood for $x^* = (x_1^*, \dots, x_m^*)'$ can be constructed by utilizing the pdf of $\nu = (\nu_2, \dots, \nu_m)'$. Assuming ε has a multivariate normal distribution with an identity covariance matrix, $\nu = (\nu_2, \dots, \nu_m)' \sim N(0, \Omega)$, where Ω is a $(m - 1) \times (m - 1)$ matrix with diagonal elements (i, i) equal to 2 and off diagonal elements (i, j) equal to 1. Given that corner solutions will occur with nonzero probability, the distribution of optimal demand will have a mixed discrete-continuous distribution with lumps of probability corresponding to regions of ε which imply corner solutions. Thus, the likelihood function will have a density component corresponding to the goods with nonzero quantities and a mass function corresponding to the corners in which some of the goods will

have zero optimal demand. The probability that n of the m goods are selected is equal to

$$\begin{aligned} P(x_i^* > 0 \text{ and } x_j^* = 0; i = 2, \dots, n \text{ and } j = n + 1, \dots, m) \\ = \int_{-\infty}^{h_m} \cdots \int_{-\infty}^{h_{n+1}} \phi(h_2, \dots, h_n, \nu_{n+1}, \dots, \nu_m | 0, \Omega) |J| d\nu_{n+1} \cdots d\nu_m, \end{aligned} \quad (3.35)$$

where $\phi(\cdot)$ is the normal density, $h_j = h_j(x^*, p)$, and J is the Jacobian,

$$J_{ij} = \frac{\partial h_{i+1}(x^*; p)}{\partial x_{j+1}^*} \quad i, j = 1, \dots, n - 1. \quad (3.36)$$

The intuition behind the likelihood function in (3.35) can be obtained from the Kuhn–Tucker conditions in (3.33) and (3.34). For goods with first-order conditions governed by (3.32), optimal demand is an implicitly defined nonlinear function of ε given by $h(\cdot)$. We use the change-of-variable theorem to derive the density of x^* (this generates the Jacobian term in (3.35)). For goods not purchased, Equation (3.34) defines a region of possible values of ν which are consistent with this specific corner solution. The probability that these goods have zero demand is calculated by integrating the normal distribution of ν over the appropriate region.

If there are only corner solutions with one good chosen, this model collapses to a standard choice model. The probability that only one good is chosen is given by

$$P(x_j^* = 0, j = 2, \dots, m) = \int_{-\infty}^{h_m} \cdots \int_{-\infty}^{h_2} \phi(\nu_2, \dots, \nu_m) d\nu_2 \cdots d\nu_m. \quad (3.37)$$

Similarly, we can derive the distribution of demand for the case in which all goods are at an interior solution.

$$P(x_i^* > 0; i = 2, \dots, m) = \phi(h_2, \dots, h_m | 0, \Omega) |J|. \quad (3.38)$$

An alternative approach to dealing with the purchase of multiple units is by modeling an aggregation process, or an aggregated event. Dube (2004) assumes the existence of latent consumption events that are represented by a model of discrete choice. Observed demand is equal

to the sum of latent events, and nonlinearities are introduced through a common satiation parameter associated with aggregated linear utility. Bradlow and Rao (2000) consider demand for product bundles in which utility is defined directly for the bundle. Both models utilize a different approach to modeling satiation (or complements, as discussed below) by utilizing properties of a discrete choice model. The approach outlined above, i.e., Kim et al. (2002), represents a departure from a standard logit model by building up the model from the Kuhn–Tucker conditions.

Models of temporal satiation, or variety seeking, has been an active research area in marketing (McAlister, 1979, 1982; Feinberg et al., 1992; Kahn, 1995). Feinberg et al. (1992) consider expected changes in market share associated with marketing variables and demonstrate that this helps to develop insights into changes in variety-seeking across product classes. They extend Lattin and McAlister's (1985) first-order Markov model in which transition probabilities are expressed as functions of variety-seeking intensity, brand preference, and brand positioning. Chintagunta (1999) proposes a “Lightning Bolt” model (Roy et al., 1996) along with a hazard model to accommodate the effects of habit persistence, unobserved heterogeneity, and state dependence on household brand choice behavior. This model links brand choice and purchase timing behavior via the effect of state-dependence. As with other approaches mentioned above, these models describe variety seeking by building descriptive models of discrete choice, and do not directly deal with nonlinear utility.

3.6 Complementary Offerings

The basic notion of a complementary offering is the presence of interaction effects among the goods. In marketing, complements can take the form of similarly branded items across categories (cake mix and frosting), or a base unit and attachments. Examples of the later include razor handles and blades, electronic toothbrushes, children's dolls (e.g., Barbie) with multiple outfits and accessories, and printers and printing supplies. The base unit is often bundled with an attachment, or a small supply of attachments, for initial purchase. Additional attachments, or refills, are usually offered separately for sale. The topic of

complementary offerings has received much attention in the marketing literature, and to date, researchers have taken various approaches to modeling their demand.

Initial attempts related to complementary offerings in marketing dealt with cross-category effects using correlated error terms and correlated response coefficients. Manchanda et al. (1999) develop a multivariate probit model in a Hierarchical Bayes framework that incorporates complementarity, co-incidence, and heterogeneity as the factors that affect multi-category choice. In their model, co-incidence is captured by the correlated error structure among utility functions of different categories and their results show the extent of the relationship between categories that arises from uncontrollable and unobserved factors. The correlation coefficients between cake mix and cake frosting and detergent and softener are higher than those of other pairs. Correlated errors is also used by Li et al. (2005) to study how customer demand for multiple products evolves over time and its implications for the sequential acquisition patterns of naturally ordered products. Correlated error terms, however, only imply that utilities move together without uncovering any underlying reasons.

Correlated coefficients represent another approach to modeling complements. Ainslie and Rossi (1998) investigate the similarities of consumer sensitivity to variables such as prices across multiple categories. A hierarchical model structure is developed to account for heterogeneity across households and categories, and an explicit correlation structure in sensitivity across categories enables to measure the degree of similarity in consumer behavior. Singh et al. (2005) estimate correlations across categories for product attributes where the intrinsic utility for a brand is a function of underlying attributes, some of which are common across categories. Similarly, Hansen et al. (2006) investigate whether the tendency to buy store brand is category specific or an enduring consumer trait. The authors develop a multi-category brand choice model with a factor-analytic structure on the covariance matrix of the coefficients and find strong evidence of correlations in household preferences for store brands across categories. Erdem (1998) and Erdem and Winer (1999) investigate preference similarity across categories due to umbrella branding. Like correlations

among errors, correlations among coefficients make utilities of two different but related categories be cross-related. In many cases, these correlations are explained by underlying factors such as sociodemographics (Erdem et al., 2001), households' shopping behavior (Ainslie and Rossi, 1998) and brand preferences (Hansen et al., 2006). However, as noted earlier, correlated coefficients do not necessarily imply a complementary relationship between two categories, but indicate the consistency of consumers' purchase (or choice) behavior among multiple categories.

In some product categories, the utility of a product depends on the number of consumers who have adopted the product and the availability of complements. These effects are referred to as direct and indirect network externality, respectively (Tirole, 1988). Indirect network externality (INE) is conceptually similar to complementarity in the sense that it deals with different but closely related categories, and builds a relationship among demand categories. In marketing, INE has only been studied with aggregate-level demand data, whereas multi-category relationship such as complementarity is more focused on individual-level consumer behavior.

Basu et al. (2003) demonstrate that INE (Indirect Network Externality) effects can vary by product attributes when externality-sensitive attributes gaining more from increased availability of complementary products than any other attributes. Gupta et al. (1999) model the market-mediated interdependence between the actions of hardware manufacturers and software complementors, created by the direct dependence of consumer demand for the whole product on the actions of manufacturers as well as complementors. Nair et al. (2004) present a framework to measure empirically the size of indirect network effects in high-technology markets with competing incompatible technology standards. The utility for software, for example, depends on hardware ownership, and conditional analysis of software preferences can be modeled with standard methods. However, such an approach does not represent the joint decision well, especially when the joint decision is made simultaneously.

Economically, complements are goods with negative (compensated) cross-substitution effects, so that a price decrease in one good

generates a demand increase in the complement (see Deaton and Muellbauer (1980, Chap. 2)). One approach is to employ a quadratic utility function:

$$u(x) = \alpha'x - 0.5x'Bx, \quad (3.39)$$

where α is a vector of coefficients, and B is a symmetric and positive definite matrix of coefficients. The associated marginal utilities are

$$u'(x) = \alpha - Bx \quad (3.40)$$

and we have the restriction $\alpha - Bx > 0$ for all x for Equation (3.39) to be a valid utility function. Equation (3.40) can then be used in the Kuhn–Tucker conditions to derive demand equations.

A drawback of this approach is the proliferation of coefficients associated with the utility function in (3.39). Additional research is needed to suggest ways of imposing *a priori* constraints on the coefficients to make this approach tractable. An approach based on an indirect utility specification is due to Song and Chintagunta (2007) and Mehta (2007).

3.7 Incorporating Attributes

Understanding the drivers of utility plays a central role in marketing theory and study. Marketing's role within a firm is to guide management to offer goods and services that consumer will want to buy, and the quantification of demand for product attributes and features has occupied a prominent place in the study of consumer behavior. Researchers can take two views on how to incorporate attributes and product characteristics into models of demand.

The first is to view the product space to be of finite dimension that does not increase as the number of offerings increase. This view is consistent with many Lancasterian models of demand that attempt to characterize demand in terms of a finite number of dimensions. For offerings comprising base units and attachments, demand is characterized in terms of the basic products used to form the available bundles — e.g., a starter kit comprises of a base unit and a number of attachments, while the refill kit consists only of attachments. Utility for the available goods (x) is viewed as an aggregation of utility for the attributes, or

characteristics (c):

$$\max_x u(c = Wx) \quad \text{subject to } p'x \leq E, \quad (3.41)$$

where W is a matrix that maps the available offerings (x) into the characteristics space (c). Marginal utility is defined in terms of the characteristics, and the marginal utility of an offering is therefore a weighted sum of the marginal utilities of the characteristics, i.e., $\partial u/\partial x = (\partial u/\partial c)(\partial c/\partial x)$:

$$\frac{\partial u}{\partial x} = \left(\frac{\partial u}{\partial c} \right)' \frac{\partial c}{\partial x'} = \left(\frac{\partial u}{\partial c} \right)' W. \quad (3.42)$$

The Kuhn–Tucker conditions involving u_i are thus defined in terms of the characteristics, and analysis proceeds as above using Equations (3.29) and (3.30) (see Kim et al., 2008). Chan (2006) proposes use of a method-of-moments estimator instead of the Kuhn–Tucker conditions for characteristics model estimation.

A second approach to incorporating attributes into models of choice is to take the view that products are relatively unique, implying that the dimension of the characteristic space is much greater than the number of offerings. There is much empirical support for this view — i.e., attempts to describe demand in terms of a low dimension characteristics space usually does not work. The high-dimension view is more commonly employed in marketing analysis, and implies that parameters of a consumer’s utility function need not reside in an identifiable subspace associated with observed characteristics. Instead, analysis proceeds by investigating various projections of utility function parameters (e.g., ψ) onto the characteristics space.

The best known technique for projecting marginal utility onto the characteristics space is conjoint analysis. The projection occurs by allowing product attribute information to modify the marginal utility parameter ψ . This projection is usually accompanied with an intercept and error terms so that observed choice need not strictly adhere to a pre-determined mapping from characteristics to offerings. For discrete choice we have:

$$\begin{aligned} \max u(x) &= \psi'x \quad \text{subject to } p'x \leq E \\ \psi &= \beta'w + \varepsilon, \end{aligned} \quad (3.43)$$

where w denotes a vector of characteristics including an intercept, and β denote the part-worths.

Conjoint has received considerable attention in academics and industry. It has been widely used for measuring consumers' tradeoffs among multiattributed products (Johnson, 1974; Green and Srinivasan, 1978; Louviere and Woodworth, 1983; Green and Krieger, 1993; Carroll and Green, 1995). Haaijer et al. (1998) provide a comprehensive overview of the conjoint analysis in the context of choice models and highlight on the restrictive assumptions of using a MNL model. They propose using a MNP model that accounts for heterogeneity in the coefficients of the choice models.

3.8 Learning Models

The incorporation of dynamic structures into economic models of choice has been an active area of research in marketing for at least 10 years. Models of consumer learning have been introduced by a number of authors beginning with Erdem and Keane (1996) and extended by Keane (1997) and Erdem (1998). Bayes theorem is used to capture consumer learning in these models. Consumers are assumed to have prior beliefs about the quality levels of brand attributes that are updated through product usage experience and advertising exposure. Thus, in models of consumer learning, the characteristics matrix, W , is a random variable, and consumers are assumed to maximize expected utility subject to a budget constraint:

$$\max_x E_{\pi(W|I)}[u(c = Wx)] \quad \text{subject to } p'x \leq E, \quad (3.44)$$

where the expectation is taken with respect to what is currently known about the product characteristics. In Equation (3.44), knowledge is represented by the statistical distribution $\pi(W|I)$ where I indicates the current information set of the respondent. The distribution is assumed to evolve over time due to product trial and advertising, and according to Bayes theorem:

$$\pi(W|I_t = \{I_{\Delta t}, I_{t-1}\}) \propto \pi(I_{\Delta t}|W)\pi(W|I_{t-1}), \quad (3.45)$$

where $I_{\Delta t}$ represents new information received about the characteristics. This information is incorporated with past information I_{t-1} to

obtain the current information set I_t . Here, we assume that $I_{\Delta t}$ is independent of I_{t-1} given W — i.e., given W , the information received over time represents a set of independent and identically distributed (iid) information so that $\pi(I_{\Delta t}|W) = \pi(I_{\Delta t}|W, I_{t-1})$.

Respondents are assumed to become informed about the characteristics of products over time through trial and advertising. These characteristics may be shared among the offerings, as in the case of a new attribute introduced into the product category, or may be unique to the offering itself. In either case, models with learning replace the standard Kuhn–Tucker conditions with conditions that incorporate consumer expectations:

$$\frac{E_{\pi(u_i|I)}[u_i]}{p_i} = \frac{E_{\pi(u_j|I)}[u_j]}{p_j} \quad \text{for } x_i^* > 0, x_j^* > 0, \quad (3.46)$$

$$\frac{E_{\pi(u_i|I)}[u_i]}{p_i} > \frac{E_{\pi(u_j|I)}[u_k]}{p_k} \quad \text{for } x_i^* > 0, x_k^* = 0. \quad (3.47)$$

Model estimation proceeds by replacing the deterministic portion of marginal utility, e.g., ψ_i in Equation (3.13), with an expected value.

Consumer learning related research in marketing has focused on different methods of incorporating quantity, quality, and prices into various stochastic choice models. Erdem and Keane (1996) propose a framework that relies on a summary measure of product quality, while Erdem et al. (2003) propose a framework of current period purchase decisions impacted by inventory and future period price expectations. They model quantity and quality learning without any nonlinearity in the pricing. Iyengar et al. (2007) propose a Bayesian dual learning process that incorporates prior beliefs about quality and consumption quantity and also nonlinearity in prices. They demonstrate their framework in a wireless services industry and show that learning can result in a win–win situation for the consumer and the firm.

Mehta et al. (2004) propose a structural model to understand consumer forgetting, learning and habit persistence. They model consumer’s indirect utility to be a linear function of the brand’s overall quality evaluation and price. Also, consumers are believed to have imperfect information about true quality evaluations. Thus, the quality evaluations are constructed from two independent information sets.

They demonstrate their model for frequently purchased detergent data when consumers imperfectly recall prior quality evaluations.

Narayanan et al. (2005) propose a Bayesian learning process model that helps consumers update their prior beliefs and reduce uncertainty about the quality of a new product. They incorporate the impact of direct (perceived product quality) and indirect (influence preferences through goodwill accumulation) effects on consumer utility, and demonstrate their model for physician learning for new drugs. Their results indicate that marketing communication has an initial indirect effect followed by domination of direct effects. Bradlow et al. (2004) propose a learning based imputation model where respondents infer missing levels of product attributes in a partial conjoint profile. They demonstrate that prior information about a product category impacts consumers' imputation strategy.

3.9 Search Models

Models of search extend traditional learning models by incorporating the amount of learning, as well as the learning sequence, as part of the model specification. Respondents are assumed to possess some information about product attributes and prices, and weigh the expected benefit of learning more about the offerings against the cost of search. Search is assumed to proceed to maximize expected utility, so long as the gains from search exceed its cost.

Consider the search process for prices. Prior to the start of search, respondents have some knowledge of the range of prices they would expect to pay. This information can be represented by a prior distribution $\pi(p)$, where p is the price vector. Prior to initiating search on prices, respondents can form expectations of the maximum utility attainable without engaging in any price search behavior. Using the simple utility function in Equation (3.20), and ignoring the effect of the budgetary allotment which is constant across choice alternatives, we have:

$$\begin{aligned} m_0 &= \max E_{\pi(p)}[u(x_j, z)] \quad \text{subject to } p'x + z \leq E \\ &= \max\{\psi_j - \alpha_z E_{\pi(p)}[p_j] + \varepsilon_j\}, \end{aligned} \quad (3.48)$$

where m_0 is the maximum expected utility at time zero. Shoppers are assumed to engage in search behavior for brand k at time t if the expected gains from search exceed the cost of search, c : $E_k m_t - m_{t-1} > c$, where:

$$m_{t-1} = \max\{\psi_j - \alpha_z E_{\pi_{t-1}(p)}[p_j] + \varepsilon_j\} \quad (3.49)$$

and

$$E_k m_t = E_{\pi_{t-1}(p_k)}[\max\{\psi_j - \alpha_z E_{\pi_{t-1}(p)|p_k}[p_j] + \varepsilon_j\}]. \quad (3.50)$$

In this model, as respondents engage in search and learn about prices, these actual prices are used to evaluate maximum attainable utility. Thus, the expectations in Equations (3.49) and (3.50) are with respect to a distribution of known prices for alternatives included in the past search path, and the conditional distribution of prices given these known prices for the alternatives not yet searched. The key difference between Equations (3.49) and (3.50), where search is being considered for alternative k , is in how the price of the k th alternative is treated. In both equations, the respondent does not yet know the price of alternative k and must take an expectation with respect to its prior distribution. In Equation (3.49), this prior is the joint conditional distribution of prices given the known prices up to time $t - 1$ because of search behavior. If prices are expected to be correlated — e.g., alternative j and i are expected to be on sale together — then past search behavior will cause the conditional distribution to change over time.

In Equation (3.50), the maximum attainable utility is computed differently. For every possible price of alternative k which is not yet known, respondents are assumed to maximize their utility across the choice alternatives. For those alternatives whose prices are known because of search, the actual (observed) prices are used. For those alternatives not yet searched, the conditional distribution of prices, given searched prices and the price of alternative k are used. Expected attainable utility is then obtained by integrating over the marginal distribution of prices for the k th alternative.

All models of search are based on formulations similar to the above example, where respondents are envisioned to make hypothetical decisions that maximize their welfare knowing the outcome of the search

variable (e.g., price). The attainable utility in Equation (3.50) is greater than the attainable utility in Equation (3.49) because the maximization operator “max” is applied to each and every potential value of the search variable (p_k). In Equation (3.49), the “max” operator is applied after expectations are taken with regard to all prices. Thus, for every value of p_k , $E_k m_t \geq m_{t-1}$.

A challenge in estimating models of search is the dimension of the search path. With K alternatives, there are $K!$ possible search paths. Models of search quickly become unmanageable because of the high dimension “state-space,” and because details of the search process are typically not available. An alternative formulation, due to Weitzman (1979), avoids the need to specify the state-space when the objects of search (e.g., prices) are independently distributed. This alternative formulation is therefore applicable to situations where knowledge about the value of one item does not affect the what is known about the other items – e.g., when searching for low prices across retailers that do not collude, or for offerings that do not share important attributes.

With independently distributed prices, an expression similar to $E_k m_t - m_{t-1}$ can be used to solve for a maximal reservation value $\tilde{m}_{k,t-1}$ for each alternative k such that the gains to search are exactly equal to the cost of search. The reservation value is then compared to the current “sure thing” reward from not searching to determine if search should continue. The “sure thing” reward is defined as in Equation (3.49), excluding the alternatives for which price is not known for certain. As search progresses, the “sure thing” reward increases, and search is less likely to occur because expected gains ($E_k m_t - m_{t-1}$) are smaller. Optimally, search terminates whenever the “sure thing” reward exceeds the reservation value $\tilde{m}_{k,t-1}$. The assumption of independently distributed prices avoids the high dimensional state-space of search problems by associating optimality conditions with the presence or absence of search, not the specific order of search.

Prior research on models of consumer information search has primarily dealt with incorporating different factors that impact the costs and benefits associated with the search. Moorthy et al. (1997) study

what factors affect the consumer search behavior process and how these factors interact with each other. They incorporate prior brand perceptions into the proposed search process and demonstrate an inverted U-shaped relationship between search process and experience. Putrevu and Ratchford (1997) propose a model of consumer information search behavior for grocery items. They use an MNL model where consumers choose a given store based on maximum perceived utility. Perceived utility is defined as the difference between a consumer's utility for a market basket and the cost associated with the search that varies from store to store. They demonstrate that the search for information about buying groceries is related to the perceived benefits and costs associated with the search process as predicted by theory.

Ratchford et al. (2003) propose a model where consumers maximize the difference between the utility gained and the cost associated with the search process. The cost for search consists of the time involved in search incorporated as opportunity cost and cognitive costs. Zwick et al. (2003) use a model of sequential search behavior where consumers go through a staged decision process where they either accept the current alternative or continue to search and pay a fixed cost or recall an attribute (with a certain probability) that has already been inspected.

Mehta et al. (2003) propose an econometric model where consumers are aware of the price distributions and engage in price search to reduce the uncertainty about prices. Consumers are assumed to make trade-offs between the expected utility associated with extensive price search and the cost involved with the search. The cost involved in the search results in consumers only considering a subset of the available brands. These costs and the associated choices are modeled probabilistically. Using scanner panel data on detergents they demonstrate that consumers incur significant search costs for figuring out the posted prices of the brands and in-store display and feature ads reduce search costs but do not change quality perceptions. Finally, Dellaert and Haubl (2004) discuss a series of related articles and future directions in economics and psychology of consumer search process.

3.10 Forward Looking Choice Behavior and Dynamic Considerations

The static choice problem is one in which the consumer chooses among alternatives given the current environment. In the language of dynamic models, the choice is made given the current “state” of the system and choice is a special case of a dynamic decision rule which maps the current state to an action, $a_t = d(s_t)$, where s_t is the vector of state variables. In the classic static choice problem, the state consists of the current values of marketing variables (the 4P’s) including price and advertising of various sorts. Implicitly, there is an assumption that utility is derived by immediate consumption of the chosen good. At the end of each period, the system is restarted and the choice problem is reposed. In this situation, there are no dynamic considerations at all. This is true even if prices or other marketing variables follow non iid or predictable time paths. Nontrivial dynamics occur in those situations in which the choice decision affects either future utility or constraints imposed on consumption, and requires that decision-makers be forward-looking in the sense of having a nonzero discount rate on future utility.

In a dynamic discrete choice model, consumers make choices at any point in time $t = 1, \dots, T$ to maximize a future stream of expected discounted marginal utility. That is, choose $a_t = d(s_t)$ to maximize:

$$E \left[\sum_{t=1}^T \delta^{t-1} (u_{a_t}(t) + \varepsilon_{a_t}) | s_0 \right], \quad (3.51)$$

where $u_{a_t}(t)$ denotes the marginal utility of the k th choice alternative at time t , a_t indicates which choice option is taken at time t . The parameter δ is a discount factor that down weights future values of marginal utility, and s_t is the set of all factors relevant to the decision at time t . $\varepsilon_{k,t}, k = 1, \dots, J$ are the usual choice specific errors that are revealed in the period in which the choice decision is made. Thus, s_t includes the choice errors, $s_t = (x_t, \varepsilon_t)$. The problem specified in (3.51) reduces to the static choice problem if $s_t = \varepsilon_t$ and marginal utilities are time invariant. In most dynamic choice models, the state vector evolves according to a Markov Process.

$$p(s_{t+1} | s_t, a_t). \quad (3.52)$$

This means that Bellman's Principle of Optimality can be used to characterize solutions to the problem of maximizing (3.51) subject to (3.52). Associated with each alternative is a "value" function which is the solution to the following equation:

$$\begin{aligned} V_k(s_t) &= u_k(t) + \delta E[V(s_{t+1})|s_t] \\ &= u_k(t) + \delta \int V(s_{t+1})p(s_{t+1}|s_t, a_t = k) ds_{t+1}. \end{aligned} \quad (3.53)$$

Bellman's Principle of Optimality implies that the solution to the decision problem is a time-invariant decision rule which is a function of the current state only. In many dynamic applications, the choice errors are assumed to be an iid extreme value. If we assume that the evolution of the state vector is independent of the choice errors, then we can integrate out the choice errors and express the problem in terms of the x_t subvector of the state vector. In this case, the choice probabilities assume the standard MNL form except that instead of net marginal or indirect utility, we have V .

$$\Pr(k|x_t) = \frac{\exp[V_k(x_t)]}{\sum_j \exp[V_j(x_t)]} \quad (3.54)$$

and

$$\begin{aligned} V_k(x) &= u_k + \delta V(x) \\ V(x) &= E[\max(V_1(x) + \varepsilon_1, \dots, V_J(x) + \varepsilon_J)]. \end{aligned} \quad (3.55)$$

In this way, we can think of a dynamic model as simply suggesting additional variables to be included in the choice probabilities as well as a source of possible nonlinearities in the indirect utility for each alternative. Computing the value function, V , can be a very challenging numerical problem, especially in problems for which the state space is large. Rust (1987) and Keane and Wolpin (1994) discuss this problem and offer solutions. However, the problem of estimating dynamic discrete choice models is beyond the scope of this survey and remains the topic of much current research.

A classic example of how dynamic considerations can enter the choice model is the case of dynamic learning models. If consumers are uncertain about the utility afforded by a set of products, then

the state variables are enlarged to include not only current values of marketing variables but current beliefs about utility (usually summarized by the parameters of the prior distribution over marginal utility). Bayes theorem provides a “law of motion” for how state vector can be updated, namely how current beliefs will transform into next periods beliefs given current period actions. These learning models incorporate aspects of sequential statistical decision making in which the consumer has a motive to experiment with brands whose current expected marginal utility are not the maximum.

Erdem and Keane (1996) provide an illustration in marketing. Here, information is valued as the pre-posterior expectation of utility, that is, the prior expected value of utility after acquiring information via consumption. Hitsch (2006) considers a similar problem from the firm side. Here the problem is whether a new product should be introduced given information available at the time of product launch as well as when the product should be removed from the market after launch. The option value of the upside of a product means that products will be launched that have a prior expected profit of less than zero, rationalizing the high rate of introduction and failure of new products.

We should emphasize that a great deal of the current literature on learning models uses a static model of learning. The decision maker does learn about choice options but only considers maximization of single period utility and does not consider the consequences of current choices for the state of knowledge in the future. That is, the sequential motive for experimentation is absent from these models. Either static or true dynamic learning models have stationary decision rules (under suitable regularity conditions) but result in nonstationary consumer behavior. As consumers learn more about the marginal utility of the choice alternatives, we should eventually see less “variety-seeking” and more predictable patterns of choice. For this reason, these models are probably not applicable to mature product categories. However, the lack of truly long-run panels in marketing makes it difficult to fit these models and draw definitive conclusions on their suitability for the observed data.

Another way in which dynamic considerations can change a choice model is in the situation of durable and/or storable products. In these situations, there is a natural distinction between the purchase of a good

and consumption. In the case of durable goods, the purchase of the good yields a stream of future utility. The packaged goods that fascinate marketers are not strictly durable but are storable. Here there can be a separation between the purchase occasion and the consumption occasion.

For durable goods, the consumer must consider the future evolution of prices. Consider the case of a good that, to a first approximation, will only be purchased once. If prices are declining at a rate greater than the discount factor, the consumer will wait for purchase of the good. For goods that deliver utility only in one period, but are storable, the consumer can speculate by holding inventory in anticipation of future temporary price reductions or sales. Here the constraint that inventory must be greater than zero is important. That is, the consumer cannot borrow to sustain his current utility. Erdem et al. (2003) and Hendel and Nevo (2006) consider this problem with various (and different) ways of dealing with the inventory non-negativity constraint. Erdem et al. (2003) impose a “stock-out” cost function, while Hendel and Nevo avoid this problem by assuming that the marginal utility is infinite at zero. The Hendel and Nevo assumption is convenient but not realistic. The idea that a consumer would pay an infinite amount for a bottle of ketchup if he stocks out is not plausible. In general, the dynamic choice problem for storable goods involves a state vector which includes the state of inventories of each of the brands in question as well as information necessary to predict the future evolution of prices. Both Erdem et al. (2003) and Hendel and Nevo (2006) abstract from this problem, using a one-dimensional summary of inventories. However, whether or not there is a purchase in the product category will depend on consumers views of the price process as well as current inventories. Erdem et al. (2003), Hendel and Nevo (2006), and Hartmann (2006) point out that the dynamic considerations mean that the price elasticity of demand is different in the short and long run and is not invariant to the price process.

In (3.51) utility is separable across time periods. That is, current period utility is only affected by current period actions. Dynamic considerations can affect choice models if the utility is not separable. The most prominent example of this in marketing is the well-documented

state dependence in demand. Here current utility is affected by previous consumption decisions. Typically, there is a form of inertia in which consumers value a product higher if it was consumed in the past. This sort of inertia can be created by switching costs, including psychological costs. The demand model can still be a static choice model, but the firm problem of setting marketing actions involves dynamic consideration. The firm must reckon with two essential economic forces — the value of acquiring customers by enticing them to purchase your product versus the “harvesting” of value from current brand loyal customers. This can be analyzed as a Markov Decision problem with a stationary price equilibrium (see Dube et al., 2006).

A related problem is the situation in which there are network effects in demand. That is, when the utility of consuming a product depends on the adoption of the product by others. This poses a nontrivial dynamic problem to the firm that can influence the stock of adopted customers by its marketing policy. Both the state dependent demand and network effects problem can be further complicated by assuming that customers are forward-looking and anticipate the marketing policies of the firm.

Finally, dynamic considerations can arise for the consumer when the firm has in place a loyalty or reward program that makes the current price dependent on the past stream of purchases (see Hartmann and Viard (2007)). Here the consumer’s current choices will influence future prices so that the state vector must be enlarged to include some summary of past purchase behavior.

The literature on dynamic choice models has, for the most part, used simple linear functions for the period by period utility. This allows the choice problem to be isolated and quantity information ignored. While this might be appropriate for durable goods where only one unit is purchased, a more realistic dynamic demand model will be required for dynamic problems of quantity and purchase incidence.

3.11 Supply-side Models

Marketing is concerned with setting the 4 P’s (product, price, promotion, and place) based on insight into consumers’ reactions so that return on investment is maximized, or other objectives are met. Thus,

joint observations of output (y), in the form of realized demand, and particular values of the 4 P's, are likely to be related through managerial actions. This relationship is ignored by models that simply regress realized demand on marketing variables (x). Regressing y on x presupposes that there is nothing to be learned about model parameters from studying the distribution of x . This situation is technically referred to as x being exogenously determined.

If marketing variables are set by managers with insight into their effectiveness (e.g., β in demand model), the effectiveness becomes a common factor resulting in the observed x and y values. In such a situation, simply regressing y on x results in misleading inferences and actions because the explanatory variables (x) are not independently determined. The solution to this problem requires prior knowledge, with the most direct form being the manager's decision protocol. Unfortunately, this information is not usually available. However, thinking about likely decision protocols is a useful exercise to guide modeling decisions and to understand the value of currently available solutions.

A common solution to the inference problem caused by endogenously determined covariates is to use instrumental variables. Useful instruments are correlated with the marketing variables under study but exogenously determined. For such variables to exist, one has to assume that some of the variation in the marketing variables is due to exogenous factors. In the absence of any exogenous variation in the explanatory variables, useful instruments do not exist.

Regressing marketing variables onto instruments partials the observed variation in the marketing variable into exogenous and endogenous parts. The variation in x predicted by the exogenous instrument is exogenous by assumption, providing the basis for consistent inferences. Villas-Boas and Winer (1999) apply this idea to the analysis of household panel data about brand choices. Household i 's (indirect) utility from brand j at time t is

$$U_{ijt} = x'_{ijt}\beta + \xi_{jt} + \varepsilon_{ijt}, \quad (3.56)$$

where ξ_{jt} is a brand and time specific demand shock and ε_{ijt} is the random part of utility and thus exogenous (i.e., not even the respondent knows ε_{ijt} up to the very point of purchase). The brand and time

specific demand shocks may be motivated by omitted covariates, such as TV advertising. If managers setting x_{ijt} have knowledge about $\{\xi_{jt}\}$ beforehand, which can influence the choice of x_{ijt} , the composite error term $(\xi_{jt} + \varepsilon_{ijt})$ is correlated with x_{ijt} . Ignoring this correlation translates into inconsistent inferences about β .

Villas-Boas and Winer (1999) focus on price as a covariate and use lagged prices as instruments, i.e., current price is regressed on the price of last period:

$$p_{jt} = \alpha_{j0} + \alpha_{j1}p_{j,t-1} + \eta_{jt}. \quad (3.57)$$

An instrumental variables approach typically substitutes predicted values of price, $(\hat{\alpha}_{j0} + \hat{\alpha}_{j1}P_{j,t-1})$, for observed price, and analysis proceeds with this substitution for the observed data. Instead of this limited information approach, Villas-Boas and Winer pursue a full information approach based on the joint likelihood for x and y . They assume a joint normal distribution for (ξ_{jt}, η_{jt}) and estimate the correlation between the demand shock and the error term in the instrumental variable regression. This correlation can be motivated by a decomposition of η_{jt} into an exogenous part and a part that is a deterministic, linear function of the demand shock ξ_{jt} due to management use of advance knowledge about ξ_{jt} .

The advantage of the full information over the limited information approach is that more variation is available to estimate the price coefficient, unless η_{jt} is perfectly predicted from ξ_{jt} . Disadvantages of the full information approach are the additional distributional assumptions and the presumption of a linear relationship between η_{jt} and the demand shocks. A universal drawback of instrumental variable approaches is the requirement that instruments need to be independent of unobserved quantities *a priori*.

An alternative approach not requiring instruments for marketing variables is to write down the objective function by which the marketing variables are assumed to be set. If this objective function takes parameters that appear in the choice model as arguments, marketing variables are determined from within the system under study.

One objective that has been studied (e.g., Yang et al., 2003) is profit maximization in an oligopolistic setting. Profit maximizing prices are

a function of price elasticity which in turn is a function of parameters in (3.56), β and $\{\xi_{jt}\}$. To the extent that (anticipated) variation in these parameters potentially causes variation in prices, this variation in prices no longer is a means to learning (more) about β and $\{\xi_{jt}\}$ because of reverse causation.

This line of reasoning distinguishes between the cases where explicitly accounting for endogeneity only increases efficiency of estimates, and where it is required for consistent estimates. Specifically, incorrectly assuming exogeneity results in inconsistent estimates whenever parameters in the demand equation act as causes to the variation in the x variables. For this to be possible, these parameters must not be constants.

With the objective function linking the distribution of x variables to parameters in the demand equation, it is useful to distinguish two situations: First, the objective function only takes observed data as arguments in addition to parameters that appear in the demand equation. Second, the objective function takes additional model parameters as inputs to explain observed variation in x .

For example, managers may be maximizing profits with first-order condition

$$\pi = Ms(p - mc) \quad (3.58)$$

$$\frac{\partial \pi}{\partial p} = \frac{\partial}{\partial p} Ms(p - mc) = 0. \quad (3.59)$$

Here M is market size, s is market share, p is price, and mc is marginal costs. The (exogenous) variation in marginal costs translates into price variation occurring independent of demand shocks. Without (any) exogenous variation in marginal costs, (3.59) implies perfect confounding of prices and demand shocks causing identification problems.

If market size, market share, and marginal costs are observed, (3.59) becomes an indicator function. Given parameters that generate market shares in (3.56), (3.59) is a degenerate likelihood function for prices, i.e., only a subset of prices solve this equation. The situation where there is more than one solution for the first-order conditions (a subset larger than one), implying a multimodal likelihood, is known as the

multiple equilibria problem in economics (Bajari, 2003; Berry et al., 2004).

Given prices, (3.59) is an indicator function in the parameters that generate market share and thus constrains the parameters in Equation (3.56) to a subset of the parameter space. Usually, however, marginal costs have to be estimated from observed cost shifters Z assumed to be exogenous.

$$mc = Z'\delta + \eta. \quad (3.60)$$

The distribution on the exogenous but unobserved (by the analyst) source of cost variation η changes (3.59) from an indicator function into a nondegenerate likelihood for prices. η accounts for inconsistencies between demand, prices, and observed cost shifters given the model.

A disadvantage of this formulation from a marketing point of view is that it presupposes optimal behavior of managers. Managers are assumed to know the demand equation, marginal costs, and how to act accordingly. Thus, the modeling by definition cannot help them make better decisions.

An alternative approach is to use supply side models that allow for imperfect manager knowledge and/or imply heuristic decisions about setting x variables. Manchanda et al. (2004) use a descriptive, empirically motivated model relating levels of detailing to parameters in their demand equation. Thus they extract information from the joint distribution of marketing variables and demand, leaving room for recommendations to improve the current setting of x variables. Models that formalize imperfect manager knowledge about response parameters is discussed in Otter et al. (2008).

4

Beyond Economics

Economic models assume the existence of a scale-valued measure of utility. Economic theory is relatively silent about how this utility arises, and behavioral decision theory (BDT) researchers have been quick to demonstrate that utility is partially determined by the context of choice and consumption. Results from the BDT literature indicate the presence of (a) a similarity effect (Tversky, 1972) produced by adding a similar alternative to a choice set; (b) an attraction effect (Huber et al., 1982) that arises from introducing a dominated alternative; and (c) a compromise effect (Simonson, 1989) from adding an intermediate alternative within a characteristics space. These results indicate that utility is constructed and affected by the choice context (Bettman et al., 1998). Thus, choice models in marketing need to either control for contextual factors, or explicitly model the formation of utility, for choice model to be predictively valid.

4.1 Modeling Contextual Effects as Adjustments to Standard Models

A simple approach to dealing with behavioral aspects of choice is to modify economic choice models to reflect aspects of choice violations.

An early example of this is the correlated probit model, where similarity effects are reflected through correlations of random utility errors. These models do not formally motivate the presence of correlated errors. Instead, they assume that brands with similarly important attributes, not represented in the deterministic specification of the model, will give rise to a correlated set of errors. Thus, these models can be viewed as a reduced-form specification of a more fundamental underlying process. The probit model, which assumes errors are distributed multivariate normal, can flexibly accommodate the presence of similarity and departures from IIA.

The formation of consideration sets has also been an area where behavioral researchers have pointed out that the assumption of the existence of utility is in question. In most product categories, the number of available choice alternatives is much too great to assume respondents engage in active consideration of all choice offerings. If a respondent does not consider an alternative for purchase, either because it lacks important features or because of costs, it is doubtful that they would take the time to form a scale valued utility for the offering.

Gilbride and Allenby (2004) develop an economic choice model with screening rules, where choice is made from among alternatives with attributes that pass a test for inclusion. Choice probabilities are expressed in terms of utility maximization over a restricted set of brands:

$$p_i = \Pr(x_i^* > 0) = \Pr\left(\frac{\psi_{i,t}}{p_{i,t}} > \frac{\psi_{k,t}}{p_{k,t}} \text{ for any } k \text{ such that } I(x_k, \gamma) = 1\right), \quad (4.1)$$

where $I(x_k, \gamma)$ is an indicator function equal to one if the function is true. An example is $I(x_k, \gamma) = I(x_k > \gamma)$, where the value of x for the k th alternative must be greater than the threshold value γ . This formulation can be expanded to include conjunctive and disjunctive decision rules by modifying the indicator function to be $\prod_m I(x_{km} > \gamma_m)$ and $\sum_m I(x_{km} > \gamma_m)$, respectively.

Gilbride and Allenby (2006) also propose a screening rule model based on the economic value of search. This model examines the trade-off between cognitive effort and expected utility and assumes that

cognitive effort is minimized by consumers by maximizing the number of screening attributes.

4.2 Modeling Contextual Effects with Model Hierarchies

Contextual consumption effects give rise to preferences that are either determined from the preferences of others, or from contextual factors such as the assortment of alternatives available for purchase. Within a specific context, standard economic models can be used to measure consumer preferences and sensitivity to variables such as prices. Across contexts, these factors are assumed to change, and their study can be approached with either the use of hierarchical models, or the use of models of utility and preference formation.

Hierarchical models have become popular in marketing during the last 15 years as modern Bayesian methods have been developed. The computational arm of these methods is a simulation-based estimation procedure known as Markov chain Monte Carlo (MCMC). A detailed discussion of these methods can be found in Rossi et al. (2005). These methods are uniquely tailored to estimating hierarchical models, where parameters of individual-level models are related through a set of common parameters, known as hyper-parameters. This upper-level model typically contains across-respondent error known as “unobserved heterogeneity.”

To illustrate, let β_h generically denote the parameters of an arbitrary choice model, where h indexes the respondent. Within a given context, these parameters are assumed to adequately reflect aspects of marginal utility and various budgetary constraints. If these parameters change across contexts, then one approach to studying contextual effects is with the specification:

$$\beta_h = \Delta z_h + \xi; \quad \xi \sim \text{Normal}(0, V_\beta), \quad (4.2)$$

where z_h are a vector of covariates that describe contextual factors of the respondent, and Δ is a matrix of coefficients that associate variation of these contextual variables with model parameters. Traditionally, z_h has reflected demographic variables (i.e., factors describing the

respondent), but it can also be used to describe factors describing the choice environment.

Liu et al. (2008) employ a hierarchical model to investigate the “level-effect” in conjoint analysis. The level-effect is an increase in the estimated importance of attribute-levels in conjoint as the number of intermediate levels of an attribute increases. A psychological explanation of the level effect can be derived from the Range-Frequency theory of Parducci (1965) where the relative spacing of attribute-levels is thought to affect judgment (see also Cooke et al., 2004), and Krumhansl’s (1978) Distance–Density model where importance is related to the denseness of the attribute space. The formulation of Liu et al. (2008) creates a functional relationship between z_h and characteristics of the attribute-levels in a conjoint model.

Yang et al. (2002) employ a variety of hierarchical models to investigate the role of motivating conditions (z_h) in affecting brand preferences. They find evidence for a unit of analysis in marketing being individual occasions of use, with specific contextual factors, that leads to preferences.

An alternative specification for studying inter-dependent preferences, where the β_h vectors are not independently distributed, is to specify an auto-regressive relationship:

$$B = \rho WB + U, \quad (4.3)$$

where B is a matrix of coefficients with row h equal to β_h' , W is a square matrix that reflects the inter-dependent relationship among the respondents that is specified by the analyst, and U is a matrix of random-effect realizations for the respondents. In Equation (4.2), the respondent parameters β_h are explicitly related to unobserved heterogeneity, while in Equation (4.3), the relationship is implicit. The implicit relationship gives rise to a set of respondent parameters that are dependently related through W (Yang and Allenby, 2003).

4.3 Process Models

Economic models begin with the assumption that a scale value of utility exists that can be used to rank choices in terms of a respondent’s

preference order. Behavioral researchers argue that this utility is often constructed at the time of decision, and not stored in a respondent's memory. Utility construction, it is argued, is therefore context dependent.

A field of inquiry in quantitative psychology is models of choice that are derived from assumptions about latent processes that yield the observed choice outcome as their terminal state. A class of models that has seen their first applications to marketing problems in this context are sequential sampling models (Townsend and Ashby, 1983). These models assume that the decision maker proceeds by repeatedly sampling evidence from the available alternatives until a pre-specified criterion value is met and a choice occurs. Within the class of sequential sampling models one can distinguish between race and diffusion models.

Race models, or counter and accumulator models, assume that evidence in favor of the available options is accumulated in separate stores. The accumulation process stops whenever any of the stores first accumulates a pre-specified amount of evidence and the corresponding choice is made. The time at which this happens corresponds to the response time. In applications, race models often build on the assumption that evidence accrues to the stores in the form of discrete "hits" that follow independent Poisson processes.

Otter et al. 2008 develop this idea into a joint likelihood for the choice among multi-attribute alternatives and response time. For example, the likelihood of choosing alternative a from a set at time t is

$$\begin{aligned}
 p(a, t | \boldsymbol{\beta}, K, f(\cdot)) &= \frac{[\exp(\mathbf{x}'_a \boldsymbol{\beta}) f(\cdot)]^K t^{K-1} \exp(-\exp(\mathbf{x}'_a \boldsymbol{\beta}) f(\cdot) t)}{(K-1)!} \\
 &\times \prod_{j:1, \dots, J/a} \left(\int_t^\infty \frac{[\exp(\mathbf{x}'_j \boldsymbol{\beta}) f(\cdot)]^K t_j^{K-1} \exp(-\exp(\mathbf{x}'_j \boldsymbol{\beta}) f(\cdot) t_j)}{(K-1)!} dt_j \right).
 \end{aligned} \tag{4.4}$$

The decision maker is assumed to track the number of hits in favor of each alternative generated at rates $\exp(\mathbf{x}'_j \boldsymbol{\beta}) f(\cdot)$ on specific counters. As soon as any one counter reaches the threshold value K , the corresponding alternative is chosen and the race terminates. The Poisson

race model contains the multinomial logit model as a special case with $K = 1$.

They propose the threshold parameter K as a measure of respondent diligence that can be empirically distinguished from respondent tastes β . Thresholds larger than one give rise to choice probabilities that structurally depart from IIA such that the chances of choosing bad alternatives decrease disproportionately. Given the same amount of data, a larger threshold translates into more likelihood information about taste parameters than a smaller threshold.

Jointly modeling choices and response times requires modeling of constrained (processing) capacity and heterogeneous processing speeds that change over the course of the tasks m as a function of process priming through:

$$f_m(\cdot) = \exp(\delta_0 + \delta_1 m) \left(\sum_{j=1}^J \exp(\mathbf{x}'_{jm} \boldsymbol{\beta}) \right)^{-1} + \exp(\lambda_0 + \lambda_1 m). \quad (4.5)$$

Otter et al. 2008 demonstrate empirical support for the endogeneity of response times as implied by this model. Quick response times point to easy decisions where at least one of the alternatives is outstanding, and slow response times point to hard decisions where the alternatives are less or equally attractive.

Ruan et al. (2008) introduce dependence between the alternative specific Poisson counters. The statistical model is motivated from independent psychological processes underlying the valuation of individual attribute levels that have stochastic components. Realized valuations of attribute levels are integrated deterministically to the overall evidence in favor of a particular alternative. The alternative that first accrues an amount of evidence equal to the threshold required for a decision is chosen. Dependence between alternative specific attribute counters is a direct consequence of shared realizations of stochastic valuations of attribute levels. Two alternatives that are close in the space set up by the attributes thus not only achieve similar expected overall evaluations but each pair of realized overall evaluations is similar. For two identical alternatives any pair of realized overall evaluations is identical. Their model naturally handles attribute based dominance as a special case of similarity.

Diffusion models, or the related random walk models, assume that relative evidence is accumulated over time. In the special case of two alternatives, relative evidence is defined as the evidence difference or the natural log of the ratio. With two alternatives, both definitions result in a scalar that is assumed to evolve according to a continuous stochastic process. Evidence for one alternative is therefore simultaneously evidence against the other. Relative evidence has to exceed/fall below pre-specified boundaries for a choice to occur. The particular boundary reached first determines the choice outcome, and the time at which the boundary is reached, the response time.

Busemeyer and Townsend (1993) introduced decision field theory (DFT) as a framework for process based modeling of preferential choice among multi-attribute alternatives. DFT is built on the idea that relative evidence has to exceed/fall below pre-specified boundaries for a choice to occur. Thus, it can be viewed as an instance of a diffusion model. Roe et al. (2001) generalized the model to choices among more than two alternatives. DFT contains classic random utility theory as a special case and provides a unifying framework for explaining observed departures from Luce's choice axiom.

DFT assumes that the decision maker's attention fluctuates stochastically between the various attributes in M according to time dependent weights $W(t)$. Fluctuating attention causes positive correlations between relative valuations of alternatives, $V(t)$ with similar attributes, generating the similarity effect.

$$V(t) = CMW(t) + \varepsilon(t). \quad (4.6)$$

Here C is a contrast matrix. Valuations $\{V(t)\}_t$, accumulate to $P(t)$ and are subject to decay and competition.

$$P(t) = \sum_{j=0}^{t-1} S^j V(t-j) + S^t P(0). \quad (4.7)$$

The alternative that first accumulates P equal to a pre-specified amount is chosen at the time where $P(t)$ first equals this amount. Alternatively, at any exogenously determined time t , the alternative for which P is maximal is chosen.

Alternatives relatively closer in the attribute space are connected through negative feedback links, i.e., close alternatives inhibit each other. Diagonal elements in S smaller than 1 cause decay and negative off-diagonal elements in S create inhibition, i.e., negative feedback. Without any processing input $V(t)$, preferences $P(t)$ thus gradually decay to a state of indifference.

The inhibitory links produce the compromise and attraction effect. The inhibitory connection between the compromise option “in the middle” and each of the extreme options is stronger than that between the extreme options. This asymmetry induces positive correlation between the relative evaluations of the extreme options and thus makes the choice of the compromise option more likely (cf. Kivetz et al., 2004; Usher and McClelland, 2004).

In case of the attraction effect, the dominated alternative accumulates negative relative evidence, which through the inhibitory, negative link translates into a boost for the similar, but dominating option. Moreover, the dynamics of decision field theory link the size of these effects to the amount of time spent with the decision. The amount of time invested is in turn a function of the accuracy goal of the decision maker. For instance, the effect of inhibition accumulates over time, with the implication that the attraction effect only occurs in well deliberated choices.

DFT and its various extensions (e.g., Diederich, 1997) provide a rich framework for modeling preferential choice among multi-attribute alternatives. However, some specification issues and computational challenges have to be addressed for likelihood based inference to become practically feasible using this framework.

5

Concluding Thoughts

Our discussion began with properties of a simple logit model, and a series of questions that arise from its formulation. It was not clear, for example, how best to incorporate changing tastes and considered brands, how best to parameterize covariates, and how to represent choices in which multiple goods are purchased. To answer these questions, we explored the economic foundation of choice models from which answers to modeling questions can be assessed. Table 5.1 presents a summary of answers to these initial questions.

The advantage of economic theory, as with any theory, is that it provides two things: (i) a foundation for gauging the reasonableness of model extensions; and (ii) parsimony. Theory helps assess proposed model enhancements while avoiding a natural tendency to add effects and variables without fully understanding their implication. For example, a popular choice model prior to the advent of modern Bayesian methods was the “mother-logit” that allowed for alternative-specific coefficients for each covariate. The mother-logit model could have K price coefficients instead of just one. While such a model can flexibly capture a variety of demand patterns, researchers found that this extra flexibility was not required once random-effect models of respondent

Table 5.1 Modeling issues and solutions.

Modeling issue	Proposed solution
Choices are conditionally independent given p .	<ul style="list-style-type: none"> • Variety seeking models incorporate effects of habit persistence, unobserved heterogeneity, state dependence that do away with the assumption that choices are conditionally independent given p. (Section 3.5)
Choice probabilities are driven by parameters that do not change over time — i.e., either p_i is constant, or β_0 and β_p are constant.	<ul style="list-style-type: none"> • Models of state dependence, habit persistence, incorporating lag specifications of the preference parameters. (Section 3.2) • Models of consumer learning, that replace the standard Kuhn–Tucker conditions with conditions that incorporate consumer expectations. (Section 3.8)
Demand is represented by zero's and one's, indicating no-choice and choice.	<ul style="list-style-type: none"> • Associate no-choice with one of the multinomial outcomes and set its marginal utility to 1. (Section 3.2) • Inclusion of an outside good into the constrained utility maximization problem. (Section 3.2) • Inside good restricted to take on values of zero or one and remaining allotment is given to the outside good. (Section 3.2)
There is an explicit set of choice alternatives included in the analysis.	<ul style="list-style-type: none"> • Construction of consideration sets and economic choice models with screening rules. (Section 4.1) • Choice alternatives that account for contextual effects. (Section 4.2)
There is an explicit function form for covariates — i.e., the logarithm of price.	<ul style="list-style-type: none"> • Alternative specifications. (Section 3.2) • Understanding the association between multinomial logit model and economic choice model with constant marginal utility. (Section 3.1)
Some of the coefficients are unique to the choice alternatives (β_0), while others are constant across choice alternatives (β_p).	<ul style="list-style-type: none"> • Propose other error distributions and state dependence that address this issue. (Section 3.2)
The IIA property is invalid.	<ul style="list-style-type: none"> • Use probit model for discrete choice. (Sections 3.2 and 4.1) • Use nested logit models. (Section 3.2)
What about other error term distributions?	<ul style="list-style-type: none"> • Use of GEV distribution. (Section 3.2)
What about quantity?	<ul style="list-style-type: none"> • Use of multivariate normal distribution. (Section 3.2) • Use direct utility function capable of dealing with corner and interior solutions. (Section 3.5)
Why constant marginal utility?	<ul style="list-style-type: none"> • Linear indifference curves but non-constant marginal utility (marginal utility is a function of attainable utility). (Section 3.4)
What if per-unit prices decline with quantity purchased?	<ul style="list-style-type: none"> • Derive the likelihood from fundamental assumptions, without treating different brand-size combinations as different multinomial choice alternatives. Allow for a price function in the budget constraint that allows for price discounts. (Section 3.4)

heterogeneity was introduced. It seems that the mis-specification of heterogeneity lead to evidence in favor of alternative-specific coefficients that are difficult to theoretically justify.

Our experience is that theoretically grounded models work well. When they do not, the best course of action is to re-examine foundational assumptions, make changes, and to re-derive the likelihood function. We feel that the direct-utility approach articulated in this survey allows researchers to most naturally make changes to the model structure.

We have identified a number of fruitful areas for future research that begin at the boundaries of economic theory. Economic theory assumes the existence of a preference ordering and an associated scalar-valued utility measure. Thus, economic theory is best applied to individuals within familiar contexts. Contextual factors such as new choice sets and new consumption settings lead to variation in utility and preferences. This variation is a fruitful avenue for new models of behavior, as is any determinant of preference and choice.

Annotated Citations of Economic Choice Models in Marketing

Ainslie, A. and P. E. Rossi (1998), “Similarities in choice behavior across product categories.” *Marketing Science* **17**, 91–106.

A multi-category choice model is developed where household response coefficients are assumed to be dependent across category. Price, display, and feature sensitivity are found to be related to household specific factors.

Allenby, G. M. and P. Lenk (1995), “Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice.” *Journal of Business and Economic Statistics* **13**, 281–289.

The logistic normal regression model of Allenby and Lenk (1994) is used to explore the order of the brand choice process and estimate the magnitude of price, display, and feature advertising effects across multiple scanner panel datasets. The results indicate large merchandising effect sizes and brand-choice that is not zero-order process.

Allenby, G. M. and P. Lenk (1994), "Modeling household purchase behavior with logistics normal regression." *Journal of American Statistical Association* **89**, 1218–1231.

A discrete choice model with autocorrelated errors and consumer heterogeneity is proposed and applied to scanner panel ketchup data. Demonstrated substantial unobserved heterogeneity and autocorrelation in purchase behavior.

Allenby, G. M. and P. E. Rossi (1991), "Quality perceptions and asymmetric switching between brands." *Marketing Science* **10**(Summer), 185–204.

A discrete choice model with rotating indifference curves is used to induce income effects that result in superior goods being favored over inferior goods. The model has non-constant marginal utility, and is used to develop an objective measure of brand quality is related to rate of rotation.

Arora, N., G. M. Allenby, and J. L. Ginter (1998), "A hierarchical bayes model of primary and secondary demand." *Marketing Science* **17**, 29–44.

A translog utility function that allows for correlations between the no-purchase and brand choice outcomes is developed. The model is compared to specifications that treat each component independently.

Bettman, J. R., M. Frances Luce, and J. W. Payne (1998), "Constructive consumer choice processes." *Journal of Consumer Research* **25**(Dec), 187–217.

Consumer preferences are shown to be constructed in response to aspects of the choice task and other demands on the decision maker. The implication is that product preferences are not stable across a variety of environmental factors.

Blattberg, R. C. and K. J. Wisniewski (1989), "Price-induced patterns of competition." *Marketing Science* **8**(Fall), 291–309.

A model of price-tiers is developed based on heterogeneity in consumer preferences. Asymmetric switching among tiers is argued to result from differing consumer response to quality.

Bechtel, G. G. (1990), "Share-ratio estimation of the nested multinomial logit model." *Journal of Marketing Research* **27**(May), 232–237.

Similar alternatives are partitioned into groups with common among and across random utility error correlation. Alternatives within the same subset are correlated while utilities of alternatives in different subsets are independent.

Chan, T. Y. (2006), "Estimating a continuous hedonic-choice model with an application to demand for soft drinks." *RAND Journal of Economics* **37**(Summer), 1–17.

A hedonic choice model is developed for multiple-product, multiple-unit purchasing behavior. The model is used to explore patterns of substitution and complementary behavior in the soft-drink category.

Chiang, J. (1991), "A simultaneous approach to the whether, what and how much to buy questions." *Marketing Science* **10**(4), 297–315.

A generalized extreme value model is developed where a household's no-purchase option and brand-choice decisions are related to a common set of covariates and model parameters. The model is compared to specifications that assume independence across components.

Chib, S., P. B. Seetharaman, and A. Strijnev (2004), "Model of brand choice with a no-purchase option calibrated to scanner-panel data." *Journal of Marketing Research* **41**(May), 184–196.

No-purchase and brand-choice decisions are jointly modeled as a function of common parameters and other unobservables. The model is applied to scanner-panel data for cola product category, and is compared to other models with no-purchase option.

Chintagunta, P. K. (1992), "Estimating a multinomial probit model of brand choice using the method of simulated moments." *Marketing Science* **11**(Fall), 386–407.

Scanner panel data of catsup purchases is used to study purchases using the method of simulated moments in a probit model. The model is shown to relax the IIA assumption of a MNL model.

Chintagunta, P. K. (1993), "Investigating purchase incidence, brand choice and purchase quantity decisions of households." *Marketing Science* **12**(Spring), 184–208.

A model with unobserved heterogeneity and category purchase, brand choice and purchase quantity decisions is developed. The model is demonstrated using scanner panel data, where the effect-size of marketing variables are shown to be affected by the category purchase decision.

Chintagunta, P. K. (1999), "Variety seeking, purchase timing, and the "Lightning Bolt" brand choice model." *Management Science* **45**(April), 486–498.

A "Lightning Bolt" or LB model (Roy et al., 1996) is incorporated into a hazard model to accommodate the effects of habit persistence, unobserved heterogeneity and state dependence on household brand choice behavior. This model links brand choice and purchase timing behavior via the effects of state-dependence.

Daganzo, C. (1979), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. New York: Academic Press.

Non-iid errors are introduced into the random component of utility to avoid the implications of IIA.

Dube, J.-P. (2004), "Multiple discreteness and product differentiation: Demand for carbonated soft drinks." *Marketing Science* **23**(Winter), 66–81.

Latent consumption occasions, described by a multinomial choice model, are used to model demand reflecting interior solutions where more than one alternative is purchased.

Erdem, T. (1996), "A dynamic analysis of market structure based on panel data." *Marketing Science* **15**, 359–378.

Past purchases are shown to affect current utility evaluations in this choice model. Analysis of scanner panel data reveals that consumer choices reflect habit persistent across a variety of product categories.

Erdem, T. and M. P. Keane (1996), "Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets." *Marketing Science* **15**, 1–20.

A Bayesian learning model is proposed where brand choice

probabilities are assumed to depend on past usage experiences and advertising exposure. The proposed dynamic, forward-looking, model is shown to provide a superior fit to the data.

Erdem, T., S. Imai, and M. P. Keane (2003), "Brand and quantity choice dynamics under price uncertainty." *Quantitative Marketing and Economics* **1**, 5–64.

A choice model is developed that incorporates previous period price expectations and on-hand inventory. The model is demonstrated using a scanner panel ketchup dataset, where demand elasticities are shown to be affected by price expectations.

Feinberg, F. M., B. E. Kahn, and L. McAlister (1992), "Market share response when consumers seek variety." *Journal of Marketing Research* **29**(May), 227–237.

A first-order Markov model is developed where transition probabilities are functions of variety-seeking intensity, brand preference, and brand positioning. The model is used to relate expected changes in market share arising from changes in the marketing mix to changes in variety-seeking behavior.

Gensch, D. H. and W. W. Recker (1979), "The multinomial, multi-attribute logit choice model." *Journal of Marketing Research* **16**(February), 124–132.

The multinomial logit model and regression model are empirically compared using weekly demand data. The logit model is shown to result in elasticities consistent with diminishing marginal utility and other properties of demand models.

Gilbride, T. J. and G. M. Allenby (2006), "Estimating heterogeneous EBA and economic screening rule choice models." *Marketing Science* **25**, 494–509.

Two process models — elimination by aspects (EBA) and a screening rule model based on economic benefit — are compared using a commercial conjoint study.

Gilbride, T. J. and G. M. Allenby (2004), "A choice model with conjunctive, disjunctive, and compensatory screening rules." *Marketing Science* **23**(Summer), 391–406.

A parsimonious model of consideration set formation is incorporated into a discrete-choice probit model. Consideration sets

are formed heterogeneously using respondent-specific threshold values and attribute-levels.

Guadagni, P. M. and J. D. C. Little (1983), “A logit model of brand choice calibrated on scanner data.” *Marketing Science* **2**(Summer), 203–238.

A multinomial logit choice model is developed for analysis of scanner panel data. The analysis demonstrates the effect of variables such as brand loyalty, size loyalty, store promotions, shelf price and price cuts on share.

Hartmann, W. (2006), “Intertemporal effects of consumption and their implications for demand elasticity estimates.” *Quantitative Marketing and Economics* **4**, 325–349.

Current consumption events are assumed to affect future utility in this model of demand. The model is validated using data on golf pairings, where intertemporal substitution is found to be persist over multiple time periods.

Hendel, I. and A. Nevo (2006), “Measuring the implications of sales and consumer inventory behavior.” *Econometrica* **74**, 1637–1673.

A dynamic consumer choice model is developed and used to compare static versus dynamic price elasticities. Scanner panel data is used to demonstrate that static demand estimates overestimate own price elasticities and underestimate cross-price elasticities.

Hitsch, G. J. (2006), “An empirical model of optimal dynamic product launch and exit under demand uncertainty.” *Marketing Science* **25**, 25–50.

A sequential learning model is developed where decisions depend on information gathered sequentially. Observed product sales are used to learn about true product profitability.

Hruschka, Harald (2007), “Using a heterogeneous multinomial probit model with a neural net extension to model brand choice.” *Journal of Forecasting* **26**, 113–127.

A nonlinear specification of utilities is computed using a neural net model calibrated with choice data. A heterogeneous multinomial probit model incorporating a neural net allowing for heterogeneous price expectations is shown to outperform other models.

Huber, J., J. W. Payne, and C. Puto (1982), "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis." *Journal of Consumer Research* **9**(June), 90–98. Experimental evidence is developed showing that adding a dominated alternative to a choice set leads to an increase in probability of the dominating alternative.

Jedidi, K., C. F. Mela, and S. Gupta (1999), "Managing advertising and promotion for long-run profitability." *Marketing Science* **18**, 1–22.

A heteroscedastic, varying-parameter probit choice and regression quantity model is developed. Scanner panel data spanning eight years is used to show that advertising has a positive and significant effect on brand equity, while promotions have a negative effect.

Kamakura, W. A. and G. J. Russell (1989), "A probabilistic choice model for market segmentation and elasticity structure." *Journal of Marketing Research* **26**(Nov), 379–390.

A finite-mixture choice model that partitions the market into a discrete set of consumer segments is examined. The model is used to study competition between national brands and private labels in a product category.

Kamakura, W. A. and R. K. Srivastava (1984), "Predicting choice shares under conditions of brand interdependence." *Journal of Marketing Research* **21**(Nov), 420–434.

Dependence among choice alternatives in a probit model is reflected in error covariance terms that are assumed to be a function of paired attributes and attribute-levels.

Kamakura, W. A. and R. K. Srivastava (1986), "An ideal-point probabilistic choice model for heterogeneous preferences." *Marketing Science*, **5**(Summer), 199–218.

Multinomial probit model is used that accounts for substitutability among choice alternatives and allows for heterogeneity in preferences by estimating a distribution of ideal points.

Keane, M. P. (1992), "A note on identification in the multinomial probit model." *Journal of Business and Economic Statistics* **10**(April), 193–200.

Simulations experiments and choice data are used to demonstrate problems with statistical identification in multinomial probit models. The problems arise from difficulty in disentangling covariance parameters from regressor coefficients in these models.

- Keane, M. P. (1997), “Modeling heterogeneity and state dependence in consumer choice behavior.” *Journal of Business and Economic Statistics* **15**(July), 310–327.

A choice model that incorporates both heterogeneity and state dependence as sources of temporal persistence in brand choice is proposed. The model simultaneously includes heterogeneous preferences for common and unique attributes of brands, time invariant preference heterogeneity, autocorrelated varying preference shocks, and state dependence. The model demonstrates a positive but small impact of promotion induced purchase on future purchase probabilities using scanner ketchup data.

- Kim, J., G. M. Allenby, and P. E. Rossi (2002), “Modeling consumer demand for variety.” *Marketing Science* **21**, 223–228.

A choice model with interior and corner solutions derived from a utility function with decreasing marginal utility. Kuhn–Tucker conditions used to relate the observed data, with utility maximization in the specification of the likelihood.

- Kim, J., G. M. Allenby, and P. E. Rossi (2007), “Product attributes and models of multiple discreteness.” *Journal of Econometrics* **138**, 208–230.

Demand for product characteristics is examined using a choice model that allows for both interior and corner solutions derived from a utility function with decreasing marginal utility. Product attribute information is associated with satiation parameters and marginal utility of various utility functions.

- Lachaab, M., A. Ansari, K. Jedidi, and A. Trabelsi (2006), “Modeling preference evolution in discrete choice models: A bayesian state space approach.” *Quantitative Marketing and Economics* **4**, 57–81.

Preference evolution is modeled using a hierarchical Bayesian state space model of discrete choice. The proposed model allows simultaneous incorporation of multiple sources of preference and

choice dynamics. Demonstrate that incorporating time-variation in parameters is crucial and outperforms other models.

Lattin, J. M. (1987), “A model of balanced choice behavior.” *Marketing Science* **6**(Winter), 48–65.

A balanced model of choice where past behavior influences the utility of the items on subsequent purchases is investigated. The analysis uses a consumption diary panel dataset, where individuals are found to seek variety on certain product attributes while exhibiting loyalty towards other attributes.

Lattin, M. J. and L. McAlister (1985), “Using a variety-seeking model to identify substitute and complementary relationships among competing products.” *Journal of Marketing Research* **22**(Aug), 330–339.

A first-order Markov model is developed in which transition probabilities are expressed as a function of variety-seeking intensity, brand preference, and brand positioning.

Li, S., B. Sun, and R. T. Wilcox (2005), “Cross-selling sequentially ordered products: An application to consumer banking services.” *Journal of Marketing Research* **42**, 233–239.

A structural multivariate probit model is developed to model purchase co-incidence and complementarity in product offerings. The model allows for the unobserved component of utilities to be correlated by assuming that the errors follow a multivariate normal distribution, while covariates are incorporated into the model to reflect complementarity.

Manchanda, P., A. Ansari, and S. Gupta (1999), “The ‘Shopping Basket’: A model for multicategory purchase incidence decisions.” *Marketing Science* **18**, 95–114.

A multivariate probit model that incorporates complementarity, co-incidence and heterogeneity as the factors that impact multicategory choice is developed. The correlational structure among the error terms is modeled to capture co-incidence while complementarity is captured by incorporating covariates.

McAlister, L. (1979), “Choosing multiple items from a product class.” *Journal of Consumer Research* **6**(Dec), 213–224.

Attribute satiation is used to develop a model that accounts for dependence among product choices. The model is compared to random choice models, independent choice models, and balanced models.

- McAlister, L. (1982), "A dynamic attribute satiation model of variety-seeking behavior." *Journal of Consumer Research* **9**(Sept), 141–150.

Product attributes are accumulated across purchases into an inventory that is related to product satiation.

- McFadden, D. F. (1978), "Modelling the choice of residential location." In: A. Karlquist et al. (eds.): *Spatial Interaction Theory and Planning Models*, Amsterdam: North-Holland Publishing Company, pp. 75–96.

A nested multinomial logit model is proposed that avoids IIA by permitting correlations among random utilities associated with similar alternatives. A choice set is portioned into mutually exclusive subsets with correlation among random utilities with similar alternatives while utilities of alternatives in different subgroups are independent.

- Punj, G. N. and R. Staelin (1978), "The choice process for graduate business schools." *Journal of Marketing Research* **15**(Nov), 588–598.

A conditional logit model is used to examine choice based on cross-sectional data rather than longitudinal data that is traditionally used to study brand choice. The model assumes that a decision maker can rank alternatives in the order of preference and will choose from the available alternatives.

- Seetharaman, P. B. (2004), "Modeling multiple sources of state dependence in random utility models of brand choice: A distributed lag approach." *Marketing Science* **23**(2), 263–271.

A discrete choice model where the preference parameters are allowed to vary as a function of past marketing actions using on distributed lag specifications is developed. The empirical results show that a structural state dependence formulation is the most important source of source of state dependence.

Simonson, I. (1989), "Choice based on reasons: The case of attraction and compromise effects." *Journal of Consumer Research* **16**(Sept), 158–174.

Attraction (similarity) and compromise (favoring a middle option) effects are shown to be present in choice under various forms of uncertainty: consequences in the future due to current actions, and preferences in the future due to these consequences.

Steckel, J. H. and W. R. Vanhonor (1988), "A heterogeneous conditional logit model of choice." *Journal of Business and Economic Statistics* **6**(July), 391–398.

Heterogeneous non-iid errors are used to avoid the IIA property in a logit model.

Swait, J. and J. Louviere (1993), "The role of the scale parameter in the estimation and comparison of multinomial logit models." *Journal of Marketing Research* **30**(August), 305–314.

The scale of the extreme value error distribution is shown to affect coefficient estimates in the corresponding logit model. Tests are proposed for studying the influence of the scaling parameters across datasets and choice tasks.

Vilcassim, N. J. and D. Jain (1991), "Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity." *Journal of Marketing Research* **28**(Feb), 29–41.

A semi-Markov process is used to model purchase timing and brand choice. Brand choice is modeled using a finite discrete state space, and the time between purchases is modeled with a hazard function. The proposed model provides insights into the dynamics of household purchase behavior.

Wagner, U. and A. Taudes (1986), "A multivariate polya model of brand choice and purchase incidence." *Marketing Science* **5**(Summer), 219–244.

A Polya model of brand choice and purchase incidence is used to relax the assumption that choice probabilities do not change over time. The integrated stochastic model of purchase timing and brand selection incorporates the influence of marketing mix

variables, seasonality and trend, and also allows for various individual choice mechanisms.

- Wedel, M., W. A. Kamakura, W. S. DeSarbo, and F. T. Hofstede (1995), "Implications for asymmetry, nonproportionality, and heterogeneity in brand switching from piece-wise exponential mixture hazard models." *Journal of Marketing Research* **32**(Nov), 457–462.

Various behavioral effects are modeled using a flexible hazard model.

- Wedel, M. and R. Pieters (2000), "Eye fixations on advertisements and memory for brands: A model and findings." *Marketing Science* **19**(Fall), 297–312.

A binary probit type model is used to understand consumer eye fixations on print advertisements that lead to memory for advertised brands. Used eye tracking data and found systematic recency effect and small primacy effect.

- Yang, S., G. M. Allenby, and G. Fennell (2002), "Modeling variation in brand preference: The roles of objective environment and motivating conditions." *Marketing Science* **21**(Winter), 14–31.

Aspects of the objective environment and motivating conditions are incorporated into a choice model through hyper-parameters in a random-effects specification. The results indicate that the unit of analysis for choice behavior are a person-activity occasion.

- Yang, S. and G. M. Allenby (2003), "Modeling interdependent preferences." *Journal of Marketing Research* **40**, 282–294.

A Bayesian spatial autoregressive discrete choice model for understanding the interdependence of preferences among individual consumers is developed.

References

- Ainslie, A. and P. E. Rossi (1998), ‘Similarities in choice behavior across product categories’. *Marketing Science* **17**, 91–106.
- Allenby, G. M., M. J. Garratt, and P. E. Rossi (2008), ‘A model of trade-up and change in considered brands’. Working Paper, Ohio State University.
- Allenby, G. M. and P. Lenk (1994), ‘Modeling household purchase behavior with logistics normal regression’. *Journal of American Statistical Association* **89**, 1218–1231.
- Allenby, G. M. and P. Lenk (1995), ‘Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice’. *Journal of Business Economic Statistics* **13**, 281–289.
- Allenby, G. M. and P. E. Rossi (1991), ‘Quality perceptions and asymmetric switching between brands’. *Marketing Science* **10**(Summer), 185–204.
- Allenby, G. M. and P. E. Rossi (1993), ‘A marginal-predictive approach to identifying household parameters’. *Marketing Letters* **4**(3), 227–239.
- Anderson, T. W. (1971), *The Statistical Analysis of Time-Series*. New York: Wiley.

- Arora, N. G., M. Allenby, and J. L. Ginter (1998), 'A hierarchical bayes model of primary and secondary demand'. *Marketing Science* **17**, 29–44.
- Bajari, P. (2003), 'Comment'. *Quantitative Marketing and Economics* **1**, 277–283.
- Basu, A., T. Mazumdar, and S. P. Raj (2003), 'Indirect network externality effects on product attributes'. *Marketing Science* **22**, 209–221.
- Bechtel, G. G. (1990), 'Share-ratio estimation of the nested multinomial logit model'. *Journal of Marketing Research* **27**(May), 232–237.
- Berry, S., J. Levinsohn, and A. Pakes (2004), 'Automobile prices in market equilibrium'. *Econometrica* **63**(July), 841–890.
- Bettman, J. R., M. Frances Luce, and J. W. Payne (1998), 'Constructive consumer choice processes'. *Journal of Consumer Research* **25**(Dec), 187–217.
- Blattberg, R. C. and K. J. Wisniewski (1989), 'Price-induced patterns of competition'. *Marketing Science* **8**(Fall), 291–309.
- Bradlow, E. T., Y. Hu, and T.-H. Ho (2004), 'A learning-based model for imputing missing levels in partial conjoint profiles'. *Journal of Marketing Research* **41**(Nov), 361–381.
- Bradlow, E. T. and V. R. Rao (2000), 'A hierarchical bayes model for assortment choice'. *Journal of Marketing Research* **37**(May), 259–268.
- Busemeyer, J. and J. T. Townsend (1993), 'Decision field theory: A dynamic-cognitive approach to decision making'. *Psychological Review* **100**(3), 432–459.
- Carroll, D. J. and P. E. Green (1995), 'Psychometric methods in marketing research: Part I, Conjoint analysis'. *Journal of Marketing Research* **32**(Nov), 385–391.
- Chan, T. Y. (2006), 'Estimating a continuous hedonic-choice model with an application to demand for soft drinks'. *Journal of Economics* **37**(Summer), 1–17.
- Chiang, J. (1991), 'A simultaneous approach to the whether, what and how much to buy questions'. *Marketing Science* **10**(4), 297–315.
- Chib, S., P. B. Seetharaman, and A. Strijnev (2004), 'Model of brand choice with a no-purchase option calibrated to scanner-panel data'. *Journal of Marketing Research* **41**(May), 184–196.

- Chintagunta, P. K. (1992), 'Estimating a multinomial probit model of brand choice using the method of simulated moments'. *Marketing Science* **11**(Fall), 386–407.
- Chintagunta, P. K. (1993), 'Investigating purchase incidence, brand choice and purchase quantity decisions of households'. *Marketing Science* **12**(Spring), 184–208.
- Chintagunta, P. K. (1999), 'Variety seeking, purchase timing, and the "Lightning Bolt" brand choice model'. *Management Science* **45**(April), 486–498.
- Chintagunta, P. K. (2002), 'Investigating category pricing behavior in a retail chain'. *Journal of Marketing Research* **39**(May), 141–154.
- Cooke, A. D. J., C. Jasnizewski, M. Cunha Jr., S. A. Nasco, and E. De Wilde (2004), 'Stimulus context and the formation of consumer ideals'. *Journal of Consumer Research* **31**(June), 112–124.
- Daganzo, C. (1979), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. New York: Academic Press.
- Deaton, A. and J. Muellbauer (1980), *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.
- Dellaert, B. G. C. and G. Haubl (2004), 'The economics and psychology of consumer search: Process-level perspectives'. *Advances in Consumer Research* **31**, 689–691.
- Diederich, A. (1997), 'Dynamic stochastic models for decision making under time pressure'. *Journal of Mathematical Psychology* **41**, 260–274.
- Domencich, T. A. and D. McFadden (1975), *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland Publishing Co.
- Dube, J.-P. (2004), 'Multiple discreteness and product differentiation: Demand for carbonated soft drinks'. *Marketing Science* **23**(Winter), 66–81.
- Dube, J.-P., G. Hitsch, and P. E. Rossi (2006), 'Can switching costs make markets more competitive?' Working Paper, Kilts Center for Marketing, Graduate School of Business, University of Chicago.
- Erdem, T. (1996), 'A dynamic analysis of market structure based on panel data'. *Marketing Science* **15**, 359–378.
- Erdem, T. (1998), 'An empirical analysis of umbrella branding'. *Journal of Marketing Research* **35**(Aug), 339–351.

- Erdem, T., S. Imai, and M. P. Keane (2003), 'Brand and quantity choice dynamics under price uncertainty'. *Quantitative Marketing and Economics* **1**, 5–64.
- Erdem, T. and M. P. Keane (1996), 'Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets'. *Marketing Science* **15**, 1–20.
- Erdem, T., G. Mayhew, and B. Sun (2001), 'Understanding reference-price shoppers: A within- and cross-category analysis'. *Journal of Marketing Research* **38**(Nov), 445–457.
- Erdem, T. and J. Swait (2004), 'Brand credibility and its role in brand choice and consideration'. *Journal of Consumer Research* **31**(1), 191–199.
- Erdem, T. and R. S. Winer (1999), 'Econometric modeling of competition: A multi-category choice-based mapping approach'. *Journal of Econometrics* **89**(Mar/Apr), 159–175.
- Feinberg, F. M., B. E. Kahn, and L. McAlister (1992), 'Market share response when consumers seek variety'. *Journal of Marketing Research* **29**(May), 227–237.
- Fisher, R. A. (1922), 'On the mathematical foundations of theoretical statistics'. *Philosophical Transactions Roy of Society London, Series, A* **222**, 309–368.
- Gensch, D. H. and W. W. Recker (1979), 'The multinomial, multiattribute logit choice model'. *Journal of Marketing Research* **16**(Feb), 124–132.
- Gilbride, T. J. and G. M. Allenby (2004), 'A choice model with conjunctive, disjunctive, and compensatory screening rules'. *Marketing Science* **23**(Summer), 391–406.
- Gilbride, T. J. and G. M. Allenby (2006), 'Estimating heterogeneous EBA and economic screening rule choice models'. *Marketing Science* **25**, 494–509.
- Green, P. E. and A. M. Krieger (1993), 'A hybrid conjoint model with individual-level interaction estimation'. *Advances in Consumer Research* **20**, 149–154.
- Green, P. E. and V. Srinivasan (1978), 'Conjoint analysis in consumer research: Issues and outlook'. *Journal of Consumer Research* **5**(Sep), 103–123.

- Greene, W. H. (2003), *Econometric Analysis*. New Jersey: Prentice Hall.
- Guadagni, P. M. and J. D. C. Little (1983), 'A logit model of brand choice calibrated on scanner data'. *Marketing Science* **2**(Summer), 203–238.
- Gupta, S., D. C. Jain, and M. S. Sawhney (1999), 'Modeling the evolution of markets with indirect network externalities: An application to digital television'. *Marketing Science* **18**, 396–417.
- Haaijer, R., M. Wedel, M. Vriens, and T. Wansbeek (1998), 'Utility covariances and context effects in conjoint MNP models'. *Marketing Science* **17**, 236–252.
- Hanemann, M. W. (1984), 'Discrete/continuous models of consumer demand'. *Econometrica* **52**(May), 541–561.
- Hansen, K. T., V. P. Singh, and P. K. Chintagunta (2006), 'Understanding store- brand purchase behavior across categories'. *Marketing Science* **25**(Winter), 75–90.
- Hartmann, W. (2006), 'Intertemporal effects of consumption and their implications for demand elasticity estimates'. *Quantitative Marketing and Economics* **4**, 325–349.
- Hartmann, W. and V. B. Viard (2007), 'Do frequency reward programs create switching costs? A dynamic structural analysis of demand in a reward program'. *Quantitative Marketing and Economics* **6**(2), 109–137.
- Hendel, I. and A. Nevo (2006), 'Measuring the implications of sales and consumer inventory behavior'. *Econometrica* **74**, 1637–1673.
- Hitsch, G. J. (2006), 'An empirical model of optimal dynamic product launch and exit under demand uncertainty'. *Marketing Science* **25**, 25–50.
- Hruschka, H. (2007), 'Using a heterogeneous multinomial probit model with a neural net extension to model brand choice'. *Journal of Forecasting* **26**, 113–127.
- Huber, J., J. W. Payne, and C. Puto (1982), 'Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis'. *Journal of Consumer Research* **9**(June), 90–98.

- Iyengar, R., A. Ansari, and S. Gupta (2007), 'A model of consumer learning for service quality and usage'. *Journal of Marketing Research* **44**(Nov), 511–544.
- Jedidi, K., S. Jagpal, and P. Manchanda (2003), 'Measuring heterogeneous reservation prices for product bundles'. *Marketing Science* **22**(Winter), 107–130.
- Jedidi, K., C. F. Mela, and S. Gupta (1999), 'Managing advertising and promotion for long-run profitability'. *Marketing Science* **18**, 1–22.
- Jedidi, K. and J. Z. Zhang (2002), 'Augmenting conjoint analysis to estimate consumer reservation price'. *Management Science* **48**(Oct), 1350–1368.
- Johnson, N. L. and S. Kotz (1971), *Discrete Distributions: Distributions in Statistics: Discrete Distributions v. 3*. Wiley Series in Probability and Mathematical Statistics.
- Johnson, R. M. (1974), 'Trade-off analysis of consumer values'. *Journal of Marketing Research* **11**(May), 121–127.
- Kahn, B. E. (1995), 'Consumer variety-seeking among goods and services'. *Journal of Retailing and Consumer Services* **2**(Fall), 139–148.
- Kahn, B. E. (1998), 'Dynamic relationships with customers: High-variety strategies'. *Journal of Academy of Marketing Science* **74**(Winter), 491–513.
- Kamakura, W. A. and G. J. Russell (1989), 'A probabilistic choice model for market segmentation and elasticity structure'. *Journal of Marketing Research* **26**(Nov), 379–390.
- Kamakura, W. A. and R. K. Srivastava (1984), 'Predicting choice shares under conditions of brand interdependence'. *Journal of Marketing Research* **21**(Nov), 420–434.
- Kamakura, W. A. and R. K. Srivastava (1986), 'An ideal-point probabilistic choice model for heterogeneous preferences'. *Marketing Science* **5**(Summer), 199–218.
- Kannan, P. K. and G. P. Wright (1991), 'Modeling and testing structured markets: A nested logit approach'. *Marketing Science* **10**(Winter), 58–82.
- Keane, M. P. (1992), 'A note on identification in the multinomial probit model'. *Journal of Business and Economic Statistics* **10**(April), 193–200.

- Keane, M. P. (1997), ‘Modeling heterogeneity and state dependence in consumer choice behavior’. *Journal of Business and Economic Statistics* **15**(July), 310–327.
- Keane, M. P. and K. I. Wolpin (1994), ‘Solution and estimation of dynamic programming models by simulation’. *Review of Economics and Statistics* **76**(4), 648–672.
- Kim, J., G. M. Allenby, and P. E. Rossi (2002), ‘Modeling consumer demand for variety’. *Marketing Science* **21**, 223–228.
- Kivetz, R., O. Netzer, and V. “Seenu” Srinivasan (2004), ‘Extending compromise effect models to complex buying situations and other context effects’. *Journal of Marketing Research* **41**(Aug), 262–268.
- Krishnamurthi, L. (1988), ‘Conjoint models of family decision making’. *International Journal of Research in Marketing* **5**(Sep), 185–198.
- Krishnamurthi, L. and S. P. Raj (1988), ‘A model of brand choice and purchase quantity price sensitivities’. *Marketing Science* **7**(Winter), 1–20.
- Krumhansl, C. L. (1978), ‘Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density’. *Psychological Review* **85**(5), 445–463.
- Lachaab, M., A. Ansari, K. Jedidi, and A. Trabelsi (2006), ‘Modeling preference evolution in discrete choice models: A bayesian state space approach’. *Quantitative Marketing and Economics* **4**, 57–81.
- Lattin, J. M. (1987), ‘A model of balanced choice behavior’. *Marketing Science* **6**(Winter), 48–65.
- Lattin, J. M. and L. McAlister (1985), ‘Using a variety-seeking model to identify substitute and complementary relationships among competing products’. *Journal of Marketing Research* **22**(Aug), 330–339.
- Li, S., B. Sun, and R. T. Wilcox (2005), ‘Cross-selling sequentially ordered products: An application to consumer banking services’. *Journal of Marketing Research* **42**, 233–239.
- Liu, Q., A. Dean, D. Bakken, and G. M. Allenby (2008), ‘Efficient experimental designs for hyper-parameter estimation: Studying the level-effect in conjoint analysis’. Working Paper, Ohio State University.

- Liu, Q., T. Otter, and G. M. Allenby (2007), 'Investigating endogeneity bias in marketing'. *Marketing Science* **26**, 642–650.
- Louviere, J. J. and G. Woodworth (1983), 'Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data'. *Journal of Marketing Research* **20**(Nov), 350–367.
- Luce, D. R. (1959), 'On the possible psychological laws'. *Psychological Review* **66**(2), 81–95.
- Manchanda, P., A. Ansari, and S. Gupta (1999), 'The "Shopping Basket": A model for multicategory purchase incidence decisions'. *Marketing Science* **18**, 95–114.
- Manchanda, P., P. K. Chintagunta, and P. E. Rossi (2004), 'Response modeling with nonrandom marketing-mix variables'. *Journal of Marketing Research* **41**(Nov), 467–478.
- McAlister, L. (1979), 'Choosing multiple items from a product class'. *Journal of Consumer Research* **6**(Dec), 213–224.
- McAlister, L. (1982), 'A dynamic attribute satiation model of variety-seeking behavior'. *Journal of Consumer Research* **9**(Sept), 141–150.
- McFadden, D. (1974), 'The measurement of urban travel demand'. *Journal of Public Economics* **3**(Nov), 303–328.
- McFadden, D. (1978), 'Modelling the choice of residential location'. In: A. Karlquist et al. (ed.): *Spatial Interaction Theory and Planning Models*. Amsterdam: North Holland Publishing Company, pp. 75–96.
- McFadden, D. (1984), 'Econometric analysis of qualitative response models'. In: Z. Griliches and M. Intriligator (eds.): *Handbook of Econometrics*, Vol. 2. Amsterdam: North Holland, pp. 1395–1457.
- Mehta, N. (2007), 'Investigating consumers' purchase incidence and brand choice decisions across multiple product categories: A theoretical and empirical analysis'. *Marketing Science* **26**(2), 196–217.
- Mehta, N., S. Rajiv, and K. Srinivasan (2003), 'Price uncertainty and consumer search: A structural model of consideration set formation'. *Marketing Science* **22**(Winter), 58–84.
- Mehta, N., S. Rajiv, and K. Srinivasan (2004), 'Role of forgetting in memory-based choice decisions: A structural model'. *Quantitative Marketing and Economics* **2**(Jun), 107–140.

- Mela, C. F., S. Gupta, and D. R. Lehmann (1997), 'The long-term impact of promotion and advertising on consumer brand choice'. *Journal of Marketing Research* **34**(May), 248–261.
- Moorthy, S., B. T. Ratchford, and D. Talukdar (1997), 'Consumer information search revisited: Theory and empirical analysis'. *Journal of Consumer Research* **23**(Mar), 263–277.
- Nair, H., P. Chintagunta, and J.-P. Dube (2004), 'Empirical analysis of indirect network effects in the market for digital personal assistants'. *Quantitative Marketing and Economics* **2**(Mar), 23–58.
- Narayanan, S., P. Manchanda, and P. K. Chintagunta (2005), 'Temporal differences in the role of marketing communication in new product categories'. *Journal of Marketing Research* **42**(Aug), 278–290.
- Otter, T., L.-J. Kao, C.-C. Chiu, T. J. Gilbride, and G. M. Allenby (2008), 'Evaluating the effectiveness of marketing expenditures'. Working Paper, Ohio State University.
- Otter and G. M. Allenby, T. and T. van Zandt (2008), 'An integrated model of choice and response time'. *Journal of Marketing Research* **45**(Oct).
- Parducci, A. (1965), 'Category judgment: A range-frequency model'. *Psychological Review* **72**, 407–418.
- Parker, P. M. (2000), *Physioeconomics: The Basis for Long-Run Economic Growth*. Cambridge: MIT Press.
- Pudney, S. E. (1989), *Modeling Individual Choice: The Econometrics of Corners, Kinks and Holes*. Oxford: Basil Blackwell.
- Punj, G. N. and R. Staelin (1978), 'The choice process for graduate business schools'. *Journal of Marketing Research* **15**(Nov), 588–598.
- Putrevu, S. and B. T. Ratchford (1997), 'A model of search behavior with an application to grocery shopping'. *Journal of Retailing* **73**(4), 463–486.
- Ratchford, B. T., M.-S. Lee, and D. Talukdar (2003), 'The impact of the internet on information search for automobiles'. *Journal of Marketing Research* **40**(May), 193–209.
- Roe, R. M., J. R. Busemeyer, and J. T. Townsend (2001), 'Multi-alternative decision field theory: A dynamic connectionist model of decision-making'. *Psychological Review* **108**, 370–392.

- Rossi, P. E., G. M. Allenby, and R. McCulloch (2005), *Bayesian Statistics and Marketing*. New York: John Wiley and Sons.
- Roy, R., P. K. Chintagunta, and S. Haldar (1996), 'A framework for investigating habits, "The Hand of the Past," and heterogeneity in dynamic brand choice'. *Marketing Science* **15**(Summer), 280–299.
- Ruan, S., S. MacEachern, T. Otter, and A. Dean (2008), 'Dependent poisson race models and modeling dependence in conjoint choice experiments'. *Psychometrika* **73**(2), 261–288.
- Rust, J. (1987), 'Optimal replacement of GMC bus engines: An empirical model of harold zucker'. *Econometrica* **55**(5), 755–784.
- Seetharaman, P. B. (2004), 'Modeling multiple sources of state dependence in random utility models of brand choice: A distributed lag approach'. *Marketing Science* **23**(2), 263–271.
- Silk, A. J. and G. L. Urban (1978), 'Pre-test market evaluation of new package goods: A model and measurement methodology'. *Journal of Marketing Research* **15**(May), 171–191.
- Simon, C. P. and L. Blume (1994), *Mathematics for Economists*. NY: W. W. Norton & Company.
- Simonson, I. (1989), 'Choice based on reasons: The case of attraction and compromise effects'. *Journal of Consumer Research* **16**(Sept), 158–174.
- Singh, V. P., K. T. Hansen, and S. Gupta (2005), 'Modeling preferences for common attributes in multicategory brand choice'. *Journal of Marketing Research* **42**(May), 195–209.
- Song, I. and P. K. Chintagunta (2007), 'A discrete-continuous model for multicategory purchase behavior of households'. *Journal of Marketing Research* **44**, 595–612.
- Sonnier, G., A. Ainslie, and T. Otter (2007), 'Heterogeneity distributions in willingness-to-pay in choice models'. *Quantitative Marketing and Economics* **5**(Sep), 313–331.
- Steckel, J. H. and W. R. Vanhonnacker (1988), 'A heterogeneous conditional logit model of choice'. *American Statistical Association* **6**(July), 391–398.
- Swait, J. and J. Louviere (1993), 'The role of the scale parameter in the estimation and comparison of multinomial logit models'. *Journal of Marketing Research* **30**(August), 305–314.

- Tanner, M. A. and W. Wong (1987), 'The calculation of posterior distributions by data augmentation'. *Journal of American Statistical Association* **82**, 528–549.
- Tirole, J. (1988), *The Theory of Industrial Organization*. Cambridge, MA: The MIT Press.
- Townsend, J. T. and F. G. Ashby (1983), *Stochastic Modeling of Elementary Psychological Processes*. Cambridge, UK: Cambridge University Press.
- Tversky, A. (1972), 'Elimination by aspects: A theory of choice'. *Psychological Review* **79**(July), 281–299.
- Usher, M. and J. L. McClelland (2004), 'Loss aversion and inhibition in dynamical models of multi-alternative choice'. *Psychological Review* **111**, 757–769.
- Vilcassim, N. J. and D. Jain (1991), 'Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity'. *Journal of Marketing Research* **28**(Feb), 29–41.
- Villas-Boas, M. J. and R. S. Winer (1999), 'Endogeneity in brand choice models'. *Management Science* **45**(Oct), 1324–1338.
- Wagner, U. and A. Taudes (1986), 'A multivariate polya model of brand choice and purchase incidence'. *Marketing Science* **5**(Summer), 219–244.
- Wedel, M., W. A. Kamakura, W. S. DeSarbo, and F. Ter Hofstede (1995), 'Implications for asymmetry, nonproportionality, and heterogeneity in brand switching from piece-wise exponential mixture hazard models'. *Journal of Marketing Research* **32**(Nov), 457–462.
- Wedel, M. and R. Pieters (2000), 'Eye fixations on advertisements and memory for brands: A model and findings'. *Marketing Science* **19**(Fall), 297–312.
- Weitzman, M. L. (1979), 'Optimal search for the best alternative'. *Econometrica* **47**(May), 641–657.
- Yang, S. and G. M. Allenby (2003), 'Modeling interdependent preferences'. *Journal of Marketing Research* **40**, 282–294.
- Yang, S., G. M. Allenby, and G. Fennell (2002), 'Modeling variation in brand preference: The roles of objective environment and motivating conditions'. *Marketing Science* **21**(Winter), 14–31.

- Yang, S., Y. Chen, and G. M. Allenby (2003), 'Bayesian analysis of simultaneous demand and supply'. *Quantitative Marketing and Economics* **1**(Sep), 251–275.
- Zwick, R., A. Rapoport, A. King Chung Lo, and A. V. Muthukrishnan (2003), 'Consumer sequential search: Not enough or too much'. *Marketing Science* **22**(Fall), 503–519.