Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets

Tülin Erdem • Michael P. Keane
University of California at Berkeley
University of Minnesota

Abstract

We construct two models of the behavior of consumers in an environment where there is uncertainty about brand attributes. In our models, both usage experience and advertising exposure give consumers noisy signals about brand attributes. Consumers use these signals to update their expectations of brand attributes in a Bayesian manner. The two models are (1) a dynamic model with immediate utility maximization, and (2) a dynamic "forward-looking" model in which consumers maximize the expected present value of utility over a planning horizon. Given this theoretical framework, we derive from the Bayesian learning framework how brand choice probabilities depend on past usage experience and advertising exposures. We then form likelihood functions for the models and estimate them on Nielsen scanner data for detergent.

We find that the functional forms for experience and advertising effects that we derive from the Bayesian learning framework fit the data very well relative to flexible ad hoc functional forms such as exponential smoothing, and also perform better at out-of-sample prediction. Another finding is that in the context of consumer learning of product attributes, although the forward-looking model fits the data statistically better at conventional significance levels, both models produce similar parameter estimates and policy implications. Our estimates indicate that consumers are risk-averse with respect to variation in brand attributes, which discourages them from buying unfamiliar brands.

Using the estimated behavioral models, we perform various scenario evaluations to find how changes in marketing strategy affect brand choice both in the short and long run. A key finding obtained from the policy experiments is that advertising intensity has only weak short run effects, but a strong cumulative effect in the long run.

The substantive content of the paper is potentially of interest to academics in marketing, economics and decision sciences, as well as product managers, marketing research managers and analysts interested in studying the effectiveness of marketing mix strategies. Our paper will be of particular interest to those interested in the long run effects of advertising.

Note that our estimation strategy requires us to specify explicit behavioral models of consumer choice behavior, derive the implied relationships among choice probabilities, past purchases and marketing mix variables, and then estimate the behavioral parameters of each model. Such an estimation strategy is referred to as "structural" estimation, and econometric models that are based explicitly on the consumer's maximization problem and whose parameters are parameters of the consumers' utility functions or of their constraints are referred to as "structural" models.

A key benefit of the structural approach is its potential usefulness for policy evaluation. The parameters of structural models are invariant to policy, that is, they do not change due to a change in the policy. In contrast, the parameters of reduced form brand choice models are, in general, functions of marketing strategy variables (e.g., consumer response to price may depend on pricing policy). As a result, the predictions of reduced form models for the outcomes of policy experiments may be unreliable, because in making the prediction one must assume that the model parameters are unaffected by the policy change.

Since the agents in our models choose among many alternative brands, their choice probabilities take the form of higher-order integrals. We employ Monte-Carlo methods to approximate these integrals and estimate our models using simulated maximum likelihood. Estimation of the dynamic forward-looking model also requires that a dynamic programming problem be solved in order to form the likelihood function. For this we use a new approximation method based on simulation and interpolation techniques. These estimation techniques may be of interest to researchers and policy makers in many fields where dynamic choice among discrete alternatives is important, such as marketing, decision sciences, labor and health economics, and industrial organization.

(*Brand Choice; Buyer Behavior; Choice Models; Econometric Modelling; Information Processing; Advertising*)

0732-2399/96/1501/0001$01.25

1. Introduction

In the past decade, a large number of models describing consumer brand choice dynamics in frequently purchased packaged consumer goods markets have been developed in marketing. These models have tried to capture the dynamics of brand choice by *exogenously* imposing a dependence between past and current choices (e.g., Jeuland 1978, Guadagni and Little 1983, Papatla and Krishnamurthi 1992). Consequently, these models do not provide a behavioral explanation for *why* current choices should depend on past choices. Indeed, state dependence in consumer choice behavior (i.e., the dependence of current choices on past choices) may be generated by many different consumer behavior patterns and product market features. Specifying explicitly the behavioral process that underlies state dependence may enable the researcher to better understand consumer choice processes and generate better predictive models of choice behavior.

The main contribution of this paper is to specify two such explicit behavioral models of consumer choice, derive the implied relationships among current choice probabilities, past purchases and marketing mix variables, and estimate the behavioral parameters of each model using scanner data. We refer to such explicit behavioral models as "structural." Structural models are based explicitly on the consumers' maximization problem and their parameters are parameters of the consumers' utility functions or of their constraints.

In previous scanner panel data research, many dynamic consumer choice models attempted to assess the impact of the marketing mix on consumer dynamic choice behavior within the context of consumer packaged goods without explicitly modelling the structure underlying consumer behavior patterns and choice processes (e.g., Carpenter and Lehman 1985, Tellis 1988, Pedrick and Zufryden 1991). Furthermore, models that allowed for state dependence in the form of inertia (e.g., Bawa 1991) or variety-seeking and purchase reinforcement (e.g., Kahn et al. 1986) typically imposed a dependence of choice probabilities on past purchases in an ad hoc way without modelling the behavioral process by which phenomena such as purchase reinforcement and inertia materialize.

One way to generate a dependence of current on past choices in an explicit behavioral model of brand choice is to introduce the informational aspects of the market. Specifically, the existence of imperfect information or consumer uncertainty about product characteristics, combined with learning behavior, can provide a behavioral explanation for the dependence of current on past choices (Meyer and Sathi 1985, Roberts and Urban 1988). Given such a theoretical model of behavior, we can derive the implied forms of state dependence and marketing mix responses, and see if these coincide with the observed patterns in the data. Such a "structural" modelling approach may provide a deeper understanding of behavior than do "nonstructural" models that specify correlational relationships in an ad hoc way.

Informational aspects are likely to be particularly important in "turbulent" consumer good markets. In such markets, new brands and/or new brand characteristics are frequently introduced. This creates consumer uncertainty about brand attributes. Under uncertainty, past experience with brands as well as marketing mix elements may affect a consumer's information set, which in turn affects his/her current choice. The dependence of consumers' information sets on past experience generates state dependence, and leads to a type of model that we describe as "dynamic."

In such a dynamic model, consumers may simply choose the brand that gives the highest current period expected utility (as in Roberts and Urban 1988). However, consumers may recognize that current choices affect their information set which may create an incentive to try different brands in order to learn about them. Thus, in making the current choice, consumers may consider the impact of the choice on the expected present value of utility over the lifetime or a planning horizon, rather than maximizing immediate utility. We refer to such a model as a "forward-looking dynamic" model of consumer choice behavior.[1]

The purpose of this paper is to implement empirically both a dynamic structural model with immediate utility maximization, which is a variate of the model first

[1] In previous work, Eckstein et al. (1988) focused on the forward-looking nature of choice dynamics in a model of consumer learning. However, their model did not incorporate information sources other than the experience with a brand (hence ruling out consideration of marketing mix elements as information sources).

proposed by Roberts and Urban (1988), and a forward-looking dynamic structural model first proposed by Erdem (1993). In these models, past experience and advertising exposures affect current choice behavior because they reduce uncertainty about product attributes, hence raising expected utility from a brand. The forward-looking dynamic model, in which consumers make current choices to maximize long-run expected utility, nests the dynamic model with immediate utility maximization, in which consumers make current choices only to maximize current expected utility.

In the dynamic structural model with immediate utility maximization, consumers learn about different brands both via their experiences with the brands (purchase feedback) and via advertising. In addition to that, in the forward-looking dynamic structural model consumers may sample (i.e., buy) different brands exclusively to gather information about them. The expected benefit of sampling brands is that a consumer may find a brand that is superior to those s/he has tried in the past. The expected cost of sampling is the expected foregone immediate utility that results from sampling a risky unknown brand rather than a safe familiar brand. After a certain time period, consumers may have sufficient information about brands so that learning about additional brands is not optimal (i.e., expected costs may exceed expected benefits). At that point, purchases of brands for learning purposes will decline or cease, and the consumer settles down to a consideration set of familiar brands from which s/he chooses on the basis of price and/or taste fluctuations.

Thus, the forward-looking dynamic model is consistent with the Howard and Sheth (1969) conceptual framework. That is, when consumers first enter the market for a frequently purchased product, they may sample brands to gather information (extended problem solving). Eventually the purchase decision process becomes routinized (i.e., expected cost of additional search exceeds expected benefit) and consumers may buy a subset of brands repeatedly. However, the sampling activity for learning purposes may be resumed at certain times due to attribute changes, brand introductions, brand repositionings, dramatic price cuts for unfamiliar brands, etc. Indeed, given the new brand introductions and the large number of alternatives available to consumers in frequently purchased consumer goods markets, consumers may be expected to form expectations about brands and learn about them over time even long after they have first entered the market.

Our modelling approach holds its greatest merit in handling choice in turbulent categories, where consumer learning is an important component of the choice process, and in evaluating the impact of policy interventions that create turbulence, such as the introduction of a new brand. However, it should be noted that if the learning model is correct, then even in static product markets where purchase of brands for learning purposes has essentially ceased, a learning model will fit the data better than a model that ignores learning. This is so because a learning model will specify a particular form of the state dependence that differs from those implied by alternative models.

We estimate both structural models on the Nielsen detergent data. We focus on the effects of past consumption experience and advertising exposure on consumer uncertainty about brands and hence on choice. We find that the functional form of state dependence derived from the structural learning models fits the data better than an exponential smoothing model (e.g., Guadagni and Little (1983) model). The structural models enable us to explain consumer expectations formation under uncertainty, and assess long-run effects of the marketing mix (advertising, in particular). Using the estimated structural models, we are able to perform a number of policy experiments (i.e., scenario evaluations) to find how changes in marketing strategy affect brand choice in both the short- and long-run. Some of these policy experiments are only possible via the structural approach, and hence are new to the marketing literature.

A key benefit of the structural modelling approach we take in this paper is its potential usefulness for policy evaluation. Most existing brand choice models can be interpreted as reduced forms, or approximations to the reduced forms, of the type of structural models studied in this paper. In reduced forms of these structural models, the parameters capturing the dependence of current choice on past experience are *functions* of marketing strategy variables. This creates the potential for bias in the estimated outcomes of policy experiments based on reduced form models, since when using these models for policy experiments one assumes that the model

parameters are *invariant* to policy. Structural models seek to avoid this problem since their parameters are parameters of the agents' objective functions or constraints that are in fact invariant to the policy changes under consideration (Lucas 1976, Rust 1991).

2. Modelling Dynamic Brand Choice in Turbulent Markets

2.1. Consumer Dynamic Optimization Problem

We consider a general model in which a consumer i decides among J possible alternatives sequentially in each of T discrete periods of time where T is finite. Alternatives are defined to be mutually exclusive, so that if $d_{ij}(t) = 1$ indicates that alternative j is chosen by decision maker i at time t and $d_{ij}(t) = 0$ indicates otherwise, then $\Sigma\, d_{ij}(t) = 1$. The choice set J includes the brands under analysis $(1, \ldots, m)$, an "Other" brands option which includes smaller brands $(m + 1)$ and a "No Purchase" option (0). To focus on the brand choice process, we assume also that a consumer buys one unit of a brand if he/she chooses to buy any brand (i.e., we are not modelling quantity choice decisions). Finally, it is presumed that consumers are uncertain about brand attributes and that consumption experience does not provide perfect information about brand attributes.

Let $I_i(t)$ denote the information set of consumer i at time t. Associated with each choice at time t is a current period expected utility, $E[U_{ij}(t) \mid I_i(t)]$ where $E[\cdot]$ is the mathematical expectation operator. The expected utility is known to the consumer at time t but is random from the perspective of periods prior to t. The specific form of the expected utilities $E[U_{ij}(t) \mid I_i(t)]$ will be introduced in the next section. The "state" of a consumer consists of all factors that affect current expected utilities and/or the probability distribution of the future expected utilities. Hence, the state of consumer i at time t is fully characterized by the information set $I_i(t)$. The objective of the consumer i at any time $t = 0, \ldots, T$, is to maximize the expected present value of utility over the planning horizon:

$$E\left[\sum_{\tau=t}^{T} \gamma^{\tau-t} \sum_{j\in J} E[U_{ij}(\tau) \mid I_i(\tau)] d_{ij}(\tau) \mid I_i(t)\right] \tag{1}$$

where $\gamma > 0$ is the discount factor. It should be noted that the inner summation in (1) involves the expected utilities, not the utilities, even when $\tau = t$ (which represents the current purchase instance). This is because the utility to be derived from a product is not known with certainty at the instance of purchase due to consumer uncertainty about brand attributes.

Maximization of (1) is accomplished by choice of the optimal sequence of control variables $\{d_{ij}(t)\}_{j\in J}$ for $t = 0, \ldots, T$. Define the maximal expected value of the discounted lifetime expected utility for consumer i at t as

$$V_i(I_i(t), t) = \max_{(d_{ij}(t))_{j\in J}} E \times \left[\sum_{\tau=t}^{T} \gamma^{\tau-t} \sum_{j\in J} E[U_{ij}(\tau) \mid I_i(\tau)] d_{ij}(\tau) \mid I_i(t)\right]. \tag{2}$$

The value function V depends on the state at t and, since t takes values from an interval of finite length, on t itself. It can be written as

$$V_i(I_i(t), t) = \max_{j\in J} V_{ij}(I_i(t), t)), \tag{3}$$

where $V_{ij}(I(t), t)$, the alternative specific expected lifetime value function, obeys the Bellman equation (Bellman 1957)[2]:

$$\begin{aligned} V_{ij}(I_i(t), t) &= E[U_{ij}(t) \mid I_i(t)] \\ &\quad + \gamma E[V_i(I_i(t+1), t+1) \mid I_i(t), d_{ij}(t) = 1], \\ &\quad t = 0, \ldots, T-1. \end{aligned} \tag{4}$$

At time T, the alternative specific value function is simply $V_{ij}(I_i(T), T) = E[U_{ij}(T) \mid I_i(T)]$.

As seen in (4), the alternative-specific value function assumes that future choices are made optimally following any given current period decision. The expectation in (4) is taken over the distribution of $I_i(t + 1)$ conditional on $I_i(t)$ and $d_{ij}(t) = 1$. In each period t, consumers are presumed to make brand choice decisions based on the information set $I_i(t)$. Obviously, the more information a consumer has about various brands, the higher

[2] Equation (4) is called the Bellman equation (Bellman 1957) in dynamic programming (DP). The major insight of DP is that the solution of a multiperiod problem can be reduced to a recursive solution of a sequence of two period problems. Thus, the consumer dynamic optimization problem can be solved by the method of DP by recursive application of Equation (4).

the expected maximum utility s/he can achieve. Thus, $V_i(I_i(t), t)$ is increasing in the quantity of information contained in $I_i(t)$.

Equation (4) indicates that the value of choosing alternative j at time t is equal not only to the immediate expected utility associated with choosing j at time t, but also the additional expected value at time $t + 1$ of having the augmented information set $I_i(t + 1)$. Brand choice alters the information set because consumption experience provides some information about brand attributes. This illustrates the information-gathering incentive of brand choice. A consumer may choose a brand j at time t that has lower expected utility than another brand if the purchase of that brand leads to a sufficient increase in the value of information that the consumer will have at time $t + 1$. For example, consumers may wish to purchase a brand that they know little about rather than buying a familiar brand with higher immediate expected utility because they may discover that they like the unfamiliar brand better.

We should note that if $\gamma = 0$ consumers maximize immediate expected utility rather than expected utility over the planning horizon. We call the general model that describes consumers as long-run expected utility maximizers ($\gamma > 0$) "the forward-looking dynamic structural model" and the model that describes consumers as immediate expected utility maximizers ($\gamma = 0$) "the dynamic structural model with immediate utility maximization". The general model nests the dynamic model with immediate utility maximization.

2.2. Consumer Expected Utility

We consider consumer uncertainty about brand attributes by positing that consumers are imperfectly informed and hence uncertain about each brand's mean attribute levels. Furthermore, attribute levels of different units of the same brand may not be constant (i.e., there is attribute variability about some mean attribute level.) Roberts and Urban (1988) call this "inherent product variability," but because it is at least partly under a firm's control, we refer to it as "attribute variability."

The attribute variability may be expressed as

$$A_{ijt} = A_j + \xi_{ijt}. \tag{5}$$

Here A refers to the attribute which exhibits variability and which is not perfectly observable; i indexes consumers ($i = 1, \ldots, n$), j indexes brands ($j = 1, \ldots, m$); t indexes time ($t = 1, \ldots, T$); and A_j is the mean attribute level for brand j.[3] The error term associated with attribute variability (ξ) is treated as an i.i.d. random variable, with zero mean and a variance that is constant over time. Thus, Equation 5 reflects the possibility that consumers can randomly get "lemons" or "windfalls."

Purchasing and consuming a brand provides consumers with information, but consumer experience of a brand's attribute levels may differ from the actual attribute levels received.[4] Equation (6) captures these "idiosyncratic perceptions":

$$A_{Eijt} = A_{ijt} + \eta_{ijt}. \tag{6}$$

A_E stands for the attribute level that a consumer perceived, and η_{ijt} is a mean zero i.i.d. random disturbance associated with consumer i's experience of brand j's attributes at time t. Thus, Equation 6 captures the possibility that consumers may not perfectly evaluate what they received. Substituting Equation 5 into Equation 6, we derive an expression for information that a consumer obtains by purchasing and experiencing a brand:

$$A_{Eijt} = A_j + \delta_{ijt} \quad \text{where } \delta_{ijt} = \xi_{ijt} + \eta_{ijt}. \tag{7}$$

To summarize, a consumer's experience of brand attributes fluctuates around the mean brand attribute levels. The two random components of the error term associated with fluctuations in consumer experience around mean attribute levels (η and ξ) cannot be separated empirically (unless a controlled experiment is designed). Therefore, they can be combined into a single term (δ), which refers to the variability of consumer experiences with the attribute levels of brands.

[3] For notational convenience and to be consistent with our empirical analysis, we are describing here the case of one imperfectly observable attribute.

[4] There may be several reasons why consumer experiences may not provide perfect information: it may take a long time to learn about product characteristics (e.g., a consumer may discover that the detergent s/he uses takes the color off the laundry after several months of usage); the experience also may be context dependent (e.g., a consumer may observe that a detergent does not remove a particular type of stain only after the event that such a stain is present in the consumer's laundry).

We assume that a brand's utility can be adequately approximated by an additive compensatory multiattribute utility model. Theoretical justification for the multiattribute modeling of consumer preferences is provided in the literature on the Fishbein-Rosenberg class of expectancy value models (Fishbein 1967) and the new economic theory of consumer choice advanced by Lancaster (1966). A consumer's utility of consuming a brand is given by the following expression:

$$U_{ijt} = -w_p P_{ijt} + w_A A_{Eijt} - w_A r A_{Eijt}^2 + e_{ijt} \qquad (8)$$

where U_{ijt} is the utility for consumer i conditional on choice of brand j at time t; P_{ijt} is the price consumer i pays for brand j at time t; w_p is the utility weight consumers attach to price (price response coefficient); w_A is consumers' attribute weight; r is the consumer risk coefficient; and e_{ijt} is random component associated with consumer i, brand j at time t. It also should be noted that utility is a function of experienced attribute levels and not the mean attribute levels.

Equation (8) is an indirect utility function with the income term suppressed (the income terms cancel out later in the logit formulation of choice probabilities). This specification is consistent with a family of direct utility functions which render corner solutions (Hanneman 1984), which is a desirable characteristic because we assume that consumers choose only one brand in each period. It suggests that utility is linear in the perfectly observable attribute (price). Furthermore, given a strictly positive w_A, utility is concave in A_E for $r > 0$, linear in A_E for $r = 0$, and convex in A_E for $r < 0$. Thus, if there is uncertainty about A_E, the consumer is risk averse, risk neutral or risk seeking as $r > 0$, $r = 0$ or $r < 0$, respectively.

Given Equation 8, the expected utility associated with brand j is

$$\begin{aligned} &E[U_{ij} | I_i(t)] \\ &\quad = -w_p P_{jt} + w_A E[A_{Eijt} | I_i(t)] - w_A r E[A_{Eijt} | I_i(t)]^2 \\ &\qquad - w_A r E[(A_{Eijt} - E[A_{Eijt} | I_i(t)])^2] + e_{ijt}. \end{aligned} \qquad (9)$$

The expected utility of consuming brand j at time t for consumer i, given the information consumer i has at time t, is a linear function of price, a concave, linear or convex function of the expected levels of the attributes A_E, which are imperfectly observed, as $r > 0$, $r = 0$ or $r < 0$, respectively and a linear function of the perceived "variance" in these attributes. Furthermore, the stochastic component of the utility function (e) reappears in the expected utility equation because e is stochastic only from the analyst's point of view.

Before discussing the consumer expectation formation process, it should be noted that Equations (8) and (9) apply only to the brands under analysis. However, since in the forward-looking dynamic structural model the unit of analysis is the time period t, a value has to be assigned to every possible choice at time t, and this includes the "No Purchase" choice. At time t a consumer may also buy an "Other" brand that is not included in our analysis (i.e., other small brand), so we must also specify a value for the choice "Other." We simply assume the expected utilities associated with "Other" and "No Purchase" to be a constant plus a time trend plus a stochastic error component:

$$E[U_{iOt}] = U_{iOt} = \Phi_O + \Psi_O t + e_{iOt} \qquad (10a)$$

$$E[U_{iNPt}] = U_{iNPt} = \Phi_{NP} + \Psi_{NP} t + e_{iNPt}. \qquad (10b)$$

The expected utilities for "Other" and "No Purchase" are expected to be increasing in time because consumers are learning about the "Other" brands and "No-Purchase" options over time, just as they are learning about the rest of the brands under analysis. The reason we refer to learning about "No Purchase" is that we will be investigating only liquid detergent purchases in the empirical analysis, so "No Purchase" may include purchase of a powdered detergent. Since it is beyond the scope of this paper to explicitly model learning about "Other" brands and powdered detergents, we capture it in an ad hoc way by including time trends in the "Other" and "No Purchase" expected utilities.

2.3. *Consumer Learning about Brand Attributes*

We assume that consumers learn about mean brand attribute levels in a Bayesian fashion.[5] Recall from Equa-

[5] We consider the Bayesian updating mechanism as a paramorphic representation (Hoffman 1960) of consumer behavior. Thus, consumers are assumed to behave *as if* they were Bayesian updaters. Actual consumer updating may not strictly follow Bayesian rules. For example, revisions may be too conservative compared to what Bayesian updating would suggest (Phillips and Edwards 1966). One may judge whether a particular *as if* mechanism provides a reasonable represen-

tion (7) that every time brand j is purchased the consumer receives an experience signal $A_{Eijt} = A_j + \delta_{ijt}$ about the mean attribute level A_j. In order to facilitate the construction of Bayesian updating rules, we assume that the signal noise δ_{ijt} and consumers' prior on A_j are both normally distributed. Thus, letting $t = 0$ be the time period when the brand is introduced, we have

$$\delta_{ijt} \sim N(0, \sigma_\delta^2), \quad A_j \sim N(A, \sigma_\nu^2(0)) \tag{11}$$

where $\sigma_\nu^2(0)$ is the initial variance (at $t = 0$) or uncertainty about A_j. According to (11), when a new brand is introduced, a consumer's prior is that the mean attribute level of that brand (A_j) is normally distributed about the product class mean attribute level (A). Thus, letting $I_i(0)$ denote the consumer's prior information about the product class, we have $E[A_j | I_i(0)] = A$.

We further assume that advertising messages also provide information about brand attribute levels. We posit that consumers may receive one advertising message about any of the brands that advertise in each period. If S represents the content of such messages, then each message signals about the attribute levels[6] of a brand according to

$$S_{ijt} = A_j + \varsigma_{ijt} \quad \varsigma_{ijt} \sim N(0, \sigma_\varsigma^2). \tag{12}$$

In order to facilitate construction of Bayesian updating rules, the error term associated with advertising signals (ς) is also assumed to be i.i.d. normal with zero mean and a variance which is constant over time. One may interpret σ_ς^2 as an indicator of the precision of the information contained in advertising messages. The informational content of the messages is not restricted to direct information. It also may relate to the precision of the images created, for example, by advertising jingles or taglines (e.g., the "Choosy mothers choose Jif" jingle) which may create a precise brand image and successfully position the brand in the attribute space.

Consumers use information they receive over time to update their prior expectations of brand mean attribute levels. Since ς has zero mean, the expected level of the signal and the expected mean level of the attribute at time t, given the information available to a consumer at time $t - 1$, are equal, that is, $E[S_{jt} | I_i(t-1)] = E[A_j | I_i(t-1)]$. Similarly, the expected level of information obtained via experience and the expected mean level of the attribute at time t, given the information available to a consumer at time $t - 1$, are equal, that is, $E[A_{Eijt} | I_i(t-1)] = E[A_j | I_i(t-1)]$.

A consumer updates his/her expectation of the mean attribute level (A_j) using the information contained in the surprise elements of the advertising signal and experience signal s/he receives according to the Bayesian rule (DeGroot 1970)[7] as follows:

$$\begin{aligned} E[A_j | I_i(t)] = {} & E[A_j | I_i(t-1)] \\ & + D_{1ijt}\beta_{1ij}(t)(A_{Eijt} - E[A_{Eijt} | I_i(t-1) \\ & + D_{2ijt}\beta_{2ij}(t)(S_{jt} - E[S_{ijt} | I_i(t-1)]). \end{aligned} \tag{13}$$

The variable D_1 equals one if a consumer purchases brand j (experience with brand j) at t, and is zero otherwise. D_2 takes the value of one if an advertising message about brand j is received at t, and is zero otherwise. The β are Kalman gain coefficients which are functions of perceived variance and experience (or advertising) variability as follows:

$$\beta_{1ij}(t) = \frac{\sigma_{\nu ij}^2(t)}{\sigma_{\nu ij}^2(t) + \sigma_\delta^2}, \quad \beta_{2ij}(t) = \frac{\sigma_{\nu ij}^2(t)}{\sigma_{\nu ij}^2(t) + \sigma_\varsigma^2} \tag{14}$$

where $\sigma_{\nu ij}^2(t)$ is the variance of a consumer's perception of brand j's mean attribute level, given the information available to the consumer at time t. The β coefficients can be interpreted as the weights consumers attach to alternative information sources in updating their expectations about brand attribute levels. If the information gained (or image created) from experience, for example, is more precise than that obtained via advertising, consumers will attach more weight to experience

tation of the underlying mechanism using model fit and predictive criteria.

[6] The term "advertising signal" is used here within the context of signal extraction: consumers try to use the information contained in the message taking the precision of the signal (i.e., $1/\sigma_\varsigma^2$) into consideration. Thus, we are not referring to the possibility that advertising expenditures signal the imperfectly observable attribute A (e.g., quality).

[7] Equation (13) can be rewritten to show that the "posterior" expectation about the mean attribute level, $E[A_j | I_i(t)]$, is a weighted average of the "prior" expectation and the sample means for the two information sources, experience and advertising.

information than to advertising information in updating their expectations about brands. Each time $\sigma^2_{\nu ij}(t)$ is updated, the β coefficients will be updated accordingly.

The variance of a consumer's perceptions of brand j's mean attribute level, $\sigma^2_{\nu ij}(t)$, represents the variance of a consumer's perception error. Consumer's perception errors at $t = 0$ are identical and are given by $\nu_j(0) = A - A_j$. Consumer perception errors at time t are denoted by $\nu_{ij}(t)$, and we have $\nu_{ij}(t) = E[A_j | I_{ij}(t)] - A_j$. Substituting for $\nu_{ij}(t)$ into (13), one obtains the process for the evolution of the perception errors over time:

$$\nu_{ij}(t) = \nu_{ij}(t-1) + D_{1ijt}\beta_{1ij}(t)(-\nu_{ij}(t-1) + \delta_{ijt}) + D_{2ijt}\beta_{2ij}(t)(-\nu_{ij}(t-1) + \varsigma_{jt}). \tag{15}$$

The perception error variance at time t is given by (DeGroot 1970):

$$\sigma^2_{\nu ij}(t) = \frac{1}{\dfrac{1}{\sigma^2_\nu(0)} + \dfrac{\sum_{s=0}^t D_{1ijs}}{\sigma^2_\delta} + \dfrac{\sum_{s=0}^t D_{2ijs}}{\sigma^2_\varsigma}}. \tag{16}$$

Equation 16 suggests that the perceived variance associated with a mean attribute level (and consequently the perceived variance of an attribute and the overall perceived variance of a brand) will be lower, ceteris paribus: (a) the more precise the advertising messages (i.e., the lower the advertising variability); (b) the more advertising signals that are received; (c) the more precise the information gained via consumption experience (i.e., the lower the experience variability of the product); and (d) the more experience a consumer has with a particular brand.

Consumer Brand Choice Probabilities. In order to obtain simple expressions for choice probabilities conditional on $I_i(t)$, we assume that the error terms e_{ijt} in Equations (9) and (10) are i.i.d. Gumbel. Define $\bar{E}[U_{ij} | I_i(t)]$ as the deterministic part of expected utility that does not depend on e_{ijt}. Then, in the dynamic structural model with immediate utility maximization the probability that brand j is chosen at time t, given the information set $I_i(t)$ is

$$P_j(I_i(t), t) = \int_\nu \frac{e^{\bar{E}[U_{ij}|I_i(t)]}}{\sum_{k=0}^J e^{\bar{E}[U_{ik}|I_i(t)]}} f(\nu) d\nu. \tag{17}$$

Equation 17 would be a standard logit choice probability if the analyst observed the consumer perception errors $\nu_{ij}(t)$. However, because we do not observe consumer perception errors, the choice probability is an integral over the distribution of possible ν. Thus, we have a logit conditional on serially correlated normal error terms, with the serial correlation pattern of the ν given by Equation 15. As we will describe below, estimation of such a model requires the use of simulation techniques to integrate out the ν.

Similarly, in the forward-looking dynamic structural model, the probability that alternative j is chosen at time t given the information set $I_i(t)$ is

$$P_j(I_i(t), t) = \int_\nu \frac{e^{\bar{E}[U_{ij}|I_i(t)] + \gamma E(V_i(I_i(t+1))|d_{ijt}=1, I_i(t))}}{\sum_{k=0}^J e^{\bar{E}[U_{ik}|I_i(t)] + \gamma E(V_i(I_i(t+1), t+1)|d_{ikt}=1, I_i(t))}} f(\nu) d\nu. \tag{18}$$

The probability that alternative j is chosen at t depends not only on the deterministic part of the expected utility obtained from alternative j at t, but also on how the choice of j affects the information set available to the consumer at time $t + 1$, which affects the time $t + 1$ expected utility. Note that the choice probabilities derived from the forward-looking dynamic expected utility maximization of consumers reduce to the probabilities obtained from the dynamic structural model with immediate utility maximization if the discount factor (γ) is zero. Thus, the dynamic structural model with immediate utility maximization is a special case of the forward-looking dynamic structural model.

We should also note that the solutions to structural models can be expressed as reduced form decision rules for consumer brand choices. The reduced form decision rules for structural models with immediate utility maximization typically take the form of a random utility model. For instance, the reduced form for our version of the Roberts-Urban model can be written as a logit with number of past purchases and past advertising exposures as explanatory variables. To see this, note that in Equation (9) the term $E[(A_{Eijt} - E[A_{Eijt} | I_i(t)])^2]$ can be decomposed into $\sigma^2_{\nu ij}(t) + \sigma^2_\delta$. Substituting for $\sigma^2_{\nu ij}(t)$ using (16) and for $E[A_{Eijt} | I_i(t)]$ using $\nu_{ij}(t) + A_j$, one obtains

$$E[U_{ij}|I_i(t)] = w_A A_j - w_A r A_j^2 - w_A r \sigma_\delta^2 - w_p P_{jt} - \frac{w_A r}{\frac{1}{\sigma_{\nu 0}^2} + \Sigma_t \frac{D_{1ijt}}{\sigma_\delta^2} + \Sigma_t \frac{D_{2ijt}}{\sigma_\varsigma^2}} - w_A r \nu_{ijt}^2 + w_A \nu_{ijt} - 2 w_A r A_j \nu_{ijt} + e_{ijt}. \quad (19)$$

This could be expressed as a logit model with alternative specific intercepts, price as a regressor, a nonlinear function of ΣD_1 and ΣD_2 as additional regressors, and a normal error component exhibiting a complex pattern of serial correlation. This reduced form decision rule could be approximated by including, for example, polynomials in ΣD_1 and ΣD_2 as regressors. It is impossible to write down the reduced form decision rule for the forward-looking dynamic structural model, since no closed form decision rule exists. However, the reduced form decision rule can be approximated by a logit model with functions of state variables as regressors. Notice that the approximate decision rules for our structural models (i.e., logits with brand intercepts, price, past experience and advertising exposure variables as regressors) have forms similar to the Guadagni-Little (GL) model (1983), except for the particular functional form with which expected utility depends on ΣD_1 and ΣD_2. In fact, the GL model can be viewed as an approximation to the reduced form decision rules that would arise from structural models similar to those described here but with the additional feature that mean attribute levels fluctuate over time, so that more recent purchases do provide better information about brands. We will estimate such a GL model to compare its fit to that of our structural models.

The crucial point of Equation (19) is that in a reduced form model, the parameters capturing state dependence will in general be functions of marketing mix variables (in this case σ_δ^2 and σ_ς^2). Thus, reduced form models may produce poor results when used to forecast the impact of changes in the marketing mix, since in making such a forecast one would assume (incorrectly) that the reduced form model parameters are invariant to changes in the marketing mix. Note that this problem will be present even if a reduced form model produces an excellent fit in calibration ("in-sample") and hold-out samples obtained from environments with a similar marketing mix.

3. Data

The data used to estimate and test the structural models discussed in §2 are scanner panel data for laundry detergent provided by A. C. Nielsen, Inc. The data set includes over 3,000 households, from two test markets in Sioux Falls, SD, and Springfield, MO. Daily data are available for a three-year period from 1986 through 1988. Sixty percent of the panel of households had a telemeter connected to their television set for the last 51 weeks in the data set, so commercial viewing data are available for that period.

3.1. The Product Category

We chose laundry detergent as the product category to model for the following reasons: (a) detergents are relatively frequently and regularly purchased products; (b) new brands were introduced during the time under investigation; (c) firms heavily advertise in this category; (d) previous research has shown detergents to be low in variety seeking (Elrod 1988). (The models we propose currently do not capture variety seeking.)

Although the data set contains information on both liquid and powdered detergent purchases, we investigated only liquid detergents, because (a) in model estimation, we restrict the number of unobservable brand attributes to one to keep the number of parameters to be estimated to a manageable number [Previous market maps of laundry detergent data revealed that there are two dimensions in the attribute space: cleansing power (or something similar to that) and whether the product is liquid or powdered (Elrod and Winer 1991)]; (b) we want to avoid multipurpose purchases of detergent (e.g., consumers may buy both liquid and powdered detergents, but for different uses), because that type of behavior may be captured incorrectly as brand switching due to price promotions or learning; and (c) there were three new liquid brand introductions as opposed to only one new powdered brand introduction during 1986–1988, so that there may be more scope for learning in the liquid category.

We use the Sioux Falls data as the calibration sample. In this market, the two market leaders were Wisk (Lever) and Tide (P&G). During 1986–1988, Dash, Surf and Cheer were introduced. We included seven brands in the analysis. The smaller brands and generics are included in an "Other" category. The last 51 weeks of the

1986–1988 period is used as the calibration sample period because commercial exposure data are only available during that period. The Springfield test market is used to assess out-of-sample fit.

3.2. Panel Member Selection Criteria

The households included in the calibration (and the hold-out) sample fulfill the following criteria: (a) They must have had a telemeter attached to their TV; (b) 80% of their total detergent purchases must have been liquid detergent purchases; (c) they must have had at least 20 liquid purchases over 1986–1988; (d) in the last 51 weeks (calibration sample period), they must have had at least 7 and at most 24[8] liquid purchases. The number of purchases required in (c) and (d) gives us a sufficiently large sequence of purchases for analysis and ensures relative homogeneity in purchase frequencies. 167 households fulfilled the above criteria in Sioux Falls and 48 households did so in Springfield.

3.3. Description of Variables Included in the Analysis

The two "exogenous" variables in the structural models are price and advertising exposure. Consumers are assumed to know mean price levels for each brand, but not when a brand will be on promotion. Although consumers are assumed to know all brand prices at time t, they must integrate over future prices to take the expectation in Equations (1), (2) and (4). Thus, we must assume a distribution for prices. We assume that prices follow the stochastic process $P_{ijt} = P_j + \omega_{ijt}$ with $\omega_{ijt} \sim N(0, \sigma^2_{\omega j})$, where P_{ijt} is the price that consumer i would pay for brand j at time t. P_j is the mean price of brand j. ω_{ijt} represents the deviation from the mean price that is revealed to the consumer at time t. It is assumed to be i.i.d. normal for convenience. The i subscript captures the fact that consumers go to different stores and therefore face different prices (we treat store choice as exogenous.)

We attempt to construct P_{ijt} by finding the price that was marked in the store visited by person i at the time of his/her visit in week t.[9] For brands other than that actually purchased by person i, the marked price must be filled in using purchases by other persons in that store on that day. There are many cases where no one bought a brand in a particular store on a particular day. We used a two step process to input these unobserved prices. First, we tried to use the average price marked for brand j in that store during week t, averaging over days when sales were observed. Second, if the whole week had no sales, we used the average price over non-promotion days for the whole sample period, adjusted for inflation. Given the price files constructed as described above, we formed the mean prices P_j for each brand j by averaging over all days and stores (weighted by store size) and calculated variances around these means to obtain the $\sigma^2_{\omega j}$.

Advertising exposure data were compiled from the commercial viewing files. If a household's TV was tuned to a channel on which a commercial was aired at any time during the full duration of the commercial, we assumed that the household was exposed to the commercial. If the household was exposed to a commercial at least once during the week, we assumed it received an advertising message during that given week. The "advertising frequency" for brand j equals the percentage of weeks during the calibration period that the average household was exposed to an ad for brand j. The advertising frequencies, mean prices and price variabilities associated with different brands in both cities are reported in Table 1.

[8] There were a few outliers with a very large or a very small number of purchases. We ignored multiple purchases of the same brand in a given week. The screening tests we conducted revealed that multiple purchases of the same brand or different brands in a given week is a rare event.

[9] We did not include coupons in the analysis. One problem associated with data on coupons is that one does not observe the availability of all brands' coupons. If coupon use is only observed for the brand actually purchased, the "after-coupon" price of brand j equals the price marked at a store minus a dummy to indicate whether brand j was bought or not times the coupon value. Thus, if "after-coupon" price is used as a covariate in a choice model, the presence of the brand dummy as a component of the covariate creates a very serious endogeneity problem. Hence, we tried to replace "the brand dummy times the coupon value" by a coupon availability proxy (calculated from data), which is exogenous. However, the goodness of fit of all the models was worse when measures for coupon availability were introduced than when they were not.

4. Model Estimation and Validation

4.1. The Estimation Procedure

In the models discussed in §2, the dependent variable is the choice of brand, and the explanatory variables ("state variables") are the number of advertising signals received (exogenous) and the past purchase history (endogenous). The parameters of the structural models are the utility weights (w_A), risk parameter (r), initial perceived variance ($\sigma_\nu^2(0)$), mean brand attributes (A_j), advertising variance (σ_ζ^2), experience variance (σ_δ^2), the intercepts and time trends for the "Other" and "No-Purchase" options (Φ_{NP}, Ψ_{NP}, Φ_O, Ψ_O), and, for the forward-looking model only, the discount factor (γ). Given the information on the dependent variables and the state variables in the scanner panel data set, these structural model parameters are estimated using the method of simulated maximum likelihood (SML).

Simulation estimation is necessary because the discrete choice probabilities needed to construct the likelihood functions for both structural models are high order integrals over the random variables e_{ijt} and $\nu_{ij}(t)$. Since e_{ijt} is assumed to be extreme value, a closed form exists for the choice probabilities conditional on ν_{ijt}. But consumers' perception errors $\nu_{ij}(t)$ are not observed by the analyst and must therefore be integrated over to form the choice probabilities. The order of integration is $T \times (J - 2)$, which in our case is $51 \times 7 = 357$.

In the simulation approach, one uses Monte Carlo methods to simulate the high order integrals that enter the likelihood function rather than evaluating them numerically (Pakes 1987, Lerman and Manski 1987, McFadden 1989, Pakes and Pollard 1989). These methods have been successfully applied to estimate discrete choice models with serially correlated error terms on panel data (Keane 1994, Elrod and Keane 1994).

Estimation of the forward-looking model poses an additional computational problem because the DP problem must be solved in order to form the likelihood function. This is an extremely computationally burdensome process, because the expected value functions (i.e., the $E[V(I_i(t), t) \mid d_{ij}(t-1), I_i(t-1)]$) are high order integrals over the distributions of the stochastic processes over which the consumers must form expectations (i.e., consumers do not know the time t realizations of prices or the advertising signals, nor do they yet know the experience signals that will result from different time $t - 1$ choices.

Very recently, Keane and Wolpin (1994) proposed a new approximation method to solve discrete choice dynamic programming problems based on simulation and interpolation. It requires one to simulate the expected value functions only at a subset of the state points. Then, the simulated expected maxima are used to fit an interpolating regression that provides fitted values for the expected maxima at any other point for which values are needed in the backwards recursion solution process described previously. We use Keane and Wolpin's method with expected attribute levels and perception error variances as regressors in the interpolating function to obtain approximate solutions for DP the problem.

A problem that must be confronted in estimation of any DP model is the choice of planning horizon T. It is not feasible to make T too large, for this increases the computational burden of the estimation. We estimate the DP problem using both $T = 100$ weeks (a planning horizon that extends 49 weeks past the end of the sample period) and $T = 51$ weeks (a planning horizon that extends only to the end of the sample period.) Our results are not sensitive to these specifications, so we report results only for the $T = 100$ specification.

A final problem that must be confronted in estimation of both the structural models and the reduced form decision rules is the initial conditions problem. We do not know the complete past histories of the consumers (in

Table 1 **Mean Prices, Standard Deviations and Advertising Frequencies**

	Sioux Falls			Springfield		
Brand	Mean Price	Price Standard Dev.	Advertising Freq.	Mean Price	Price Standard Dev.	Advertising Freq.
Dash	2.86	0.25	0.10	–	–	–
Cheer	3.71	0.31	0.17	4.01	0.31	0.27
Solo	3.92	0.37	0.00	–	–	–
Surf	3.49	0.42	0.23	3.46	0.39	0.13
Era	3.86	0.41	0.57	4.09	0.41	0.81
Wisk	3.26	0.35	0.74	3.65	0.38	0.17
Tide	3.71	0.36	0.51	4.13	0.28	0.11

terms of brand experiences and advertising exposures) at the point in time when our estimation period begins. We use the first two years of the Nielsen data to impute the past consumption and advertising experience of the households at the start of year 3 (i.e., the beginning of the calibration period).[10]

4.2. Data Analysis

We will first estimate, in our terminology, an approximate reduced form decision rule that takes the form of a multinomial logit model with past purchases and advertising exposures as explanatory variables. We call this the GL model (Guadagni and Little 1983). We estimate the GL model solely as a baseline model to judge if our structural models fit the data reasonably well. The GL model we estimate is a simple variate of the original GL model. In particular, we do not estimate a size loyalty effect; our price variable includes promotions, hence, we do not have two separate price coefficients; we also introduce an "advertising" variable, which is an exponentially weighted average of past advertising exposures. As in the case of the original GL model, the "brand loyalty" variable is also an exponentially weighted average of past purchases, which corresponds to "experience" effects in the structural models. We estimate the model by maximum likelihood. Following estimation of the GL model, we proceed to estimate the two dynamic structural models previously described.

4.3. Parameter Estimates

The parameters of the GL model are the price coefficient (w_p); the "brand loyalty" coefficient (w_E), which can be interpreted as the "Experience" coefficient; the "advertising" coefficient (w_{Ad}), which can be conceptualized as capturing the impact of past advertising exposure on consumer brand choice; the brand intercepts (a); the no-purchase and other brands constants (Φ_{NP} and Φ_O) and time trend coefficients (Ψ_{NP} and Ψ_O); and the experience and advertising smoothing constants (α_E and α_{Ad}). The parameter estimates and t-statistics are reported in Table 2.

The price coefficient is negative, whereas the brand loyalty coefficient is positive, and both coefficients are statistically significant. Thus, "past purchases" of a brand (or "Brand Loyalty" as GL defined it) increase the probability that the brand will be chosen in the current period. All the brand intercepts and smoothing constants are positive and significant. (The intercept of the first brand (Dash) is fixed at 0 for identification.) The most important result in Table 2 is that, although the advertising coefficient has the right sign (i.e., positive), it is not statistically significant. Thus, the estimated GL model suggests that advertising does not affect brand choice.

We turn now to the structural model estimates. In estimation of the forward-looking dynamic model we fixed the weekly discount factor at 0.995. This is common in applied dynamic programming research because of the difficulties often encountered in estimating

Table 2 GL Model Estimates

Parameter	Estimate	t-statistic
price coefficient ($-w_p$)	−1.077	−18.10
"brand loyalty" parameter (w_E)	3.363	53.18
advertising coefficient (w_{Ad})	0.144	0.31
brand intercepts (a_j):		
a_{Dash}	0.000	–
a_{Cheer}	1.115	8.87
a_{Solo}	0.917	7.22
a_{Surf}	1.382	14.43
a_{Era}	1.601	11.03
a_{Wisk}	1.102	6.78
a_{Tide}	1.700	12.29
"Other Brands" intercept (Φ_O)	−0.633	−2.98
"Other Brands" time trend (Ψ_O)	0.011	4.87
"No Purchase" intercept (Φ_{NP})	1.636	8.02
"No Purchase" time trend (Ψ_{NP})	0.005	1.35
"Brand Loyalty" smoothing coefficient (α_E)	0.770	50.62
advertising smoothing coefficient (α_{AD})	0.788	2.95

−LL = 7463.23 AIC = 7478.23 BIC = 7531.10

[10] There were no advertising data available for the first two years of data. Hence mean advertising exposure levels (probability of being exposed to an ad per week times number of weeks) were used to impute past advertising histories before the start of the calibration period. Once the state space at t_1 was imputed, the initial perception variance could be calculated and we could draw initial expected attribute levels from a normal distribution. It is worth noting that the expected attribute levels drawn at the beginning of the calibration period are not conditional on consumer (household) purchase histories, although they should be conditional on these histories. Therefore using the first two years of the data to impute the initial state is only an approximate solution to the initial conditions problem.

Table 3 Structural Model Estimates

Parameter	Immediate Utility Maximization[1] ($\gamma = 0$) Estimate	t-statistic	Forward-looking Dynamic Structural Model[2] ($\gamma = 0.995$) Estimate	t-statistic
price coefficient ($-w_p$)	−0.790	−12.26	−0.795	−12.31
utility weight (w_A)	28.356	1.73	34.785	1.84
risk coefficient (r)	3.625	2.08	4.171	2.25
initial variance ($\sigma_\nu^2(t)$)	0.053	4.64	0.040	4.21
mean attribute levels (A_j):				
A_{Dash}	0.049	0.74	0.040	0.74
A_{Cheer}	0.019	0.27	0.012	0.21
A_{Solo}	0.056	0.84	0.047	0.87
A_{Surf}	0.105	1.65	0.089	1.77
A_{Era}	0.137	2.41	0.120	2.64
A_{Wisk}	0.040	0.59	0.029	0.53
A_{Tide}	0.138	–	0.120	–
"Other Brands" intercept (Φ_O)	−17.657	−7.98	−17.267	−7.59
"Other Brands" time trend (Ψ_O)	0.018	8.53	0.018	8.91
"No Purchase" intercept (Φ_{NP})	−15.408	−6.99	−19.537	−8.55
"No Purchase" time trend (Ψ_{NP})	0.011	3.17	0.012	3.42
experience variability (σ_δ)	0.374	9.17	0.33	8.37
advertising variability (δ_ξ)	3.418	6.29	3.08	5.57

[1] −LL = 7312.09 AIC = 7324.09 BIC = 7384.49
[2] −LL = 7306.05 AIC = 7322.05 BIC = 7378.45

discount rates.[11] Besides the discount factor, both the structural models share a common set of parameters. These are the price coefficient (w_p); the weight attached to the (imperfectly observable) attribute (w_A); the risk coefficient (r); the initial variance ($\sigma_\nu^2(0)$); the actual mean attribute levels of the brands (A_j);[12] the intercepts and time trends in the payoff functions for "No Purchase" and "Other" (Φ_{NP}, Ψ_{NP}, Φ_O, Ψ_O); and the experience and advertising variabilities in standard deviation form (σ_δ and σ_ξ). We estimated one "latent" attribute for each brand (A_j). For identification reasons, since absolute attribute levels have no meaning, the attribute level of one brand must be fixed. Our quadratic utility function requires that attribute levels be below a certain level so that utility is increasing in the attribute level. Therefore, we set the attribute level of Tide (brand 7) to be 1 initially and updated it on each step of the optimization algorithm to ensure that it stayed in the proper range.

The parameter estimates and t-statistics of both the structural models are shown in Table 3. The results in

[11] The intertemporal factor is usually assumed to be between 0 and 1 because it is assumed to be 1/(1 + interest rate) although behaviorally this does not have to hold. In applied DP work it is often very difficult to get an estimate of γ, and in many cases researchers have been unable to obtain an estimate (possibly because of misspecification of the theoretical model). We first estimated the discount factor as a parameter, and it turned out to equal 1.001. However, the standard error of 0.02 was high from a theoretical point of view because it implies that the annual discount rate may fall in a wide range. When we estimated the model again with $\gamma = 0.995$, the parameter estimates were insensitive to it, and the log-likelihood was slightly worse (the log-likelihood improves with additional parameters estimated). Consequently, we can use the likelihood ratio as a conservative test to see if the difference in fit between the two learning models is statistically significant.

[12] A's are imperfectly observable from the consumers' point of view. They are "latent" variables to the analyst.

Table 4 Out-of-Sample Fit Test Statistics

Test Statistic	GL	Dynamic Structural Models: Immediate Utility Maximization	Dynamic Structural Models: Forward-looking
−LL	2000.69	1951.38	1952.98
AIC	2015.69	1967.38	1968.98
BIC	2059.21	2013.80	2015.40

Table 3 suggest that the parameter estimates of the two structural models are very similar. The price coefficient has the right sign and is statistically significant. Consumers attach a high weight to the "latent" attribute level which we interpret as "quality" or "cleansing" power. In both models Tide and Era are estimated to have the highest level of the latent attribute, followed next by Surf, then by Solo, Dash and Wisk, and finally by Cheer. Note that most of the estimated attribute levels are not significantly different from zero, but this is not important, because zero does not have any absolute meaning for attribute level estimates.[13]

The risk coefficient is significantly positive which indicates risk-averse behavior. This result, combined with a positive w_A, indicates that increased perceived attribute variance decreases expected utility and lowers brand choice probability. The experience and advertising variabilities (σ_δ and σ_ς, respectively) are both statistically significant, with advertising variability being much higher than experience variability. This result suggests that experience provides much more information than advertising about laundry detergents. The initial variance ($\sigma_\nu^2(0)$) is significantly positive, but small, indicating that consumers' prior about quality of new brands is concentrated around the product class mean.

[13] It is worth noting that the standard errors of the estimated attribute levels are high because one can easily scale the attribute levels up or down (along with corresponding changes in other model parameters) without substantially changing the overall results. Thus, the model has some difficulty in estimating "absolute" levels of the latent attributes. This also partly explains the high standard error on w_A. However, the estimated covariances among the estimated A's are also high, so that the differences among the attribute levels are indeed significant.

The estimates of the time trends in the "Other" and "No Purchase" payoff functions suggest that the payoff associated with other brands and no purchase are increasing over time, as expected. The constants of these payoff functions are negative, because we obtained negative utilities, given the functional form of the utility function and our parameter estimates. It should be noted that "utility" does not have any cardinal meaning in the proposed theory (only relative magnitudes are meaningful).

4.4. Goodness of Fit

Table 4 reports the log-likelihood values (LL) for the calibration sample, as well as the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). These results suggest that both structural models fit better than the GL model. This is a rather surprising result because the structural models impose a great deal of structure on the relationship between past history and current choices, while the GL model allows a very flexible exponential smoothing form for the effect of past history on current choices.

We should note that it would be simple to incorporate additional explanatory variables into the logit model used here to improve its fit. It would be very difficult to make the structural models more complex. Thus, one should not justify structural models on "fit" because they do worse in terms of "simplicity of use." The justification for structural models is the insight into processes they provide, and the policy experiments that they can be used for. Note that the GL model does not explain how brand loyalty is formed, how past purchases affect current choices or how utility changes over time. Both structural models, however, suggest that past purchases affect current choices because the experience associated with brands (information feedback) decreases perceived variance (risk) which increases expected utility. Given risk-aversion, consumers tend to buy brands with which they have had positive experience. Hence, brand loyalty is an outcome of positive use experience.

So far, we contrasted the overall fit of the GL model with the structural models. Now we consider the differences between the structural models. As previously indicated, the parameter estimates of the two models are very similar. The likelihood ratio statistic (−2 times

Table 5 Aggregate Choice Probabilities

	Dash	Cheer	Solo	Surf	Era	Wisk	Tide	Other	No Purchase
				Sioux Falls					
Actual	0.0072	0.0114	0.0075	0.0223	0.0352	0.0409	0.0491	0.0666	0.7599
GL	0.0075	0.0114	0.0074	0.0221	0.0361	0.0421	0.0502	0.0672	0.7599
I. Dynamic	0.0097	0.0113	0.0083	0.0198	0.0310	0.0484	0.0424	0.0670	0.7622
F. Dynamic	0.0100	0.0119	0.0085	0.0208	0.0317	0.0496	0.0431	0.0674	0.7570
				Springfield					
Actual	–	0.0090	–	0.0049	0.0515	0.0310	0.0613	0.0739	0.7684
GL	–	0.0063	–	0.0056	0.0520	0.0346	0.0453	0.0700	0.7862
I. Dynamic	–	0.0048	–	0.0011	0.0524	0.0306	0.0351	0.0708	0.8052
F. Dynamic	–	0.0053	–	0.0012	0.0533	0.0314	0.0354	0.0714	0.8020

the difference in the log likelihood functions) is 12.08 when the forward-looking dynamic structural model is compared to the dynamic structural model with immediate utility maximization. The critical value of χ^2 with 1 degree of freedom at $\alpha = 0.01$ is 6.64. Thus, we reject the null hypothesis that the discount factor γ is zero, and conclude that consideration of the impact of current choices on future expected utilities does improve the in-sample fit of the model. However, although statistically significant, the difference in the fit is not "quantitatively" large. One reason for this is that even if consumers are forward looking, the value associated with sampling brands as part of the information gathering process is not very high because the product category is mature and in spite of turbulence due to brand introductions consumers may not expect new brands to differ substantially from existing brands. This is indicated in the parameter estimates by the fact that the prior variance $\sigma_v^2(0)$ is small.

To test the predictive validity of the models, we predicted the choices of the 48 households in the Springfield sample. Thus, the hold-out sample consisted of different households than those in the calibration sample. We should note that predictive validity tests based upon purchases of different households from the same city, or of the same households but for different time periods, would be less stringent than tests based on a different set of individuals from a different city. Prediction for a different city is more difficult because of the great differences in the product market environments between the cities. For example, relative prices and advertising frequencies (See Table 1) differ greatly between the cities, and Dash and Solo liquid detergents were never introduced in Springfield.

In terms of the likelihood function, AIC and BIC, both of the structural models predict the choices of the households better than the simple GL model. The dynamic structural model with immediate utility maximization predicts slightly better than the forward-looking dynamic structural model, but the difference is not statistically significant. We also calculated the predicted aggregate choice probabilities for each brand both in-sample and out-of-sample. The results are reported in Table 5.

The results in Table 5 indicate that the GL model predicts the aggregate choice probabilities better both in-sample and out-of-sample. This is not surprising because the brand constants in the GL model adjust to capture the aggregate market shares.[14] Although the GL model predicts aggregate shares better, the structural models provide more insight into the determination of aggregate shares. For example, in Sioux Falls, Dash has the lowest market share. GL simply adjusts the brand constants to capture market share, and as the market share of Dash is the lowest, the brand constant is the

[14] Note that the A's of the structural models do not behave like brand intercepts. The A's are determined jointly with expectation errors and hence cannot adjust freely to reflect market shares.

lowest as well. On the other hand, structural models "explain" why the market share of Dash may be the lowest: a low advertising frequency (See Table 1), rather than a low attribute level, gives rise to a relatively high risk which, in turn, gives rise to a relatively low level of expected utility, rather than a low attribute level. (Note that according to the structural models the cleansing power of Dash is not the lowest among the set of brands.)

5. Scenario Evaluations

In this section we use the two structural models estimated in §4 to simulate how changes in the marketing mix affect consumer behavior. A key advantage of structural models is that they can be used to perform policy experiments that do not suffer from the Lucas (1976) critique of econometric policy evaluation. In estimating structural models we have estimated parameters of consumers' utility functions. These are invariant to changes in the marketing mix. Hence, under the usual assumption that we have the "true" model, we can obtain consistent estimates of the impact of changes in firm strategy on consumer choices. It is important to note that this is not true for reduced form models. The parameters of reduced form models are in general functions of marketing mix variables. An example of this can be seen in Equation (19), in which the parameters capturing state dependence are functions of the marketing mix variables (e.g., σ_ζ^2). Thus, a policy experiment may cause the parameters of a reduced form model to change, rendering its predictions of behavioral response unreliable.

As a concrete example, in a "reduced form" or "nonstructural" model like the simple GL model the link between past purchases and current choice probabilities may be a function of strategy variables like advertising frequency. An increase in advertising frequency increases the exogenous rate of information arrival, thus reducing the incentive for an agent to purchase a new brand for trial (ceteris paribus). Consequently, for example, with a higher advertising frequency a consumer may require, ceteris paribus, a deeper price cut for a new brand in order to induce trial. (It is crucial to note that "ceteris paribus" here includes holding fixed the number of advertisements already seen.)

An additional key advantage of structural models is that they can be used to perform some policy experiments that are impossible using nonstructural models. A nonstructural model can only estimate effects of policy variables that are observed by the analyst, so that they can be included as explanatory variables (e.g., price or number of advertising exposures.) In our structural models, advertising and experience variability are policy variables which affect the sequence of signals received by agents. But since these signals are not observed by the analyst, it is impossible to estimate their effect on choice in a nonstructural framework.

In order to simulate choice behavior, we used the two structural models estimated in §4 and randomly drew advertising and experience signals and prices to simulate choice outcomes. We varied the advertising frequency and advertising variability of various brands to assess the impact of such policy changes on brand choice. Figures 1 through 8 display a representative sample of our results. Each figure represents the impact of various strategy changes on the aggregate choice probabilities for "Surf".[15] We chose Surf to conduct the simulation experiments because Surf was relatively a new brand during the time period studied.

In Figures 1 through 8, the X-axis is Time in weeks. The Y-axis is the aggregate choice probabilities of Surf. Thus, the baseline (heavy lines) represents the aggregate choice probability associated with Surf in a given week prior the strategy change. The broken lines refer to the post-strategy case. We ran experiments for both structural models under two different conditions: (1) A "New Product" condition in which a brand with the same parameters (e.g., attribute level, mean price level, price variability, advertising frequency) of an actual brand like Surf was assumed to have been introduced in week 1,[16] and (2) a "Mature Product" condition in which we assess what would have happened in a hypothetical situation where Surf had pursued a different

[15] Unfortunately, the experiments assume away all possible "reactions" from competitors, because we would have to make the firms' actions endogenous in the theoretical model to be able to include competitor reactions. This is beyond the scope of this paper.

[16] Note that Surf was introduced during the 1986–1988 period, but before the first week of the calibration period.

Figure 1 Surf Increases Ad Frequency from 0.234 to 0.700

(I. Dynamic/New Product)

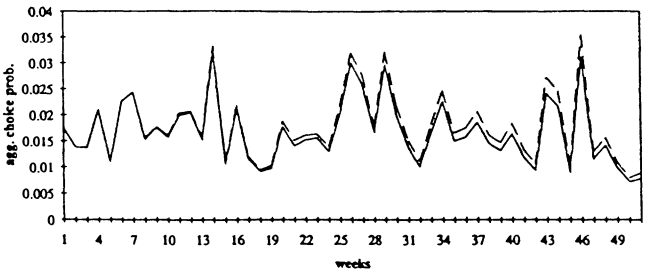


Figure 2 Surf Increases Ad Frequency from 0.234 to 0.700

(F. Dynamic/New Product)

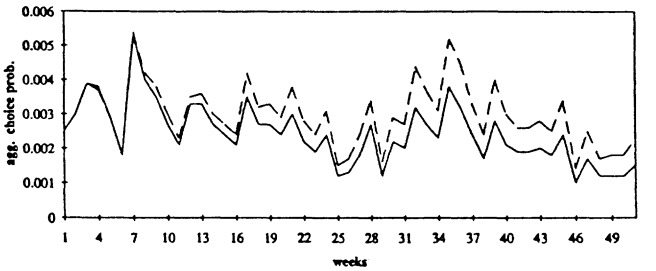


Figure 3 Surf Increases Ad Frequency from 0.234 to 0.700

(I. Dynamic/Mature Product)

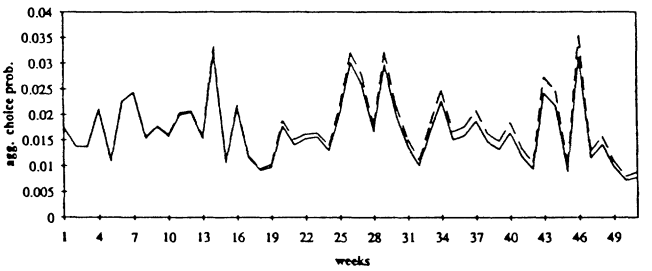
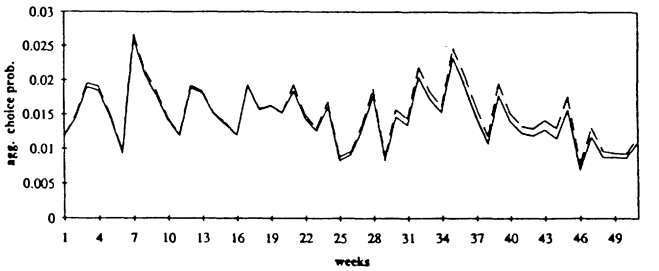


Figure 4 Surf Increases Ad Frequency from 0.234 to 0.700

(F. Dynamic/Mature Product)



marketing strategy. The results can be summarized as follows:

(1) The higher the advertising frequency, the higher the brand choice probability, and this holds more for a "new product" than a "mature product." Furthermore, although the short run effect of advertising is not large, advertising has a strong cumulative effect on choice over time as it gradually reduces the perceived riskiness of a brand. This can be seen in the results for both structural models (Figures 1 through 4).

(2) Figure 5 and 6 show that lowering advertising variability increases choice probabilities. Thus, precise advertising messages have a major positive impact on new product sales growth. In the "mature product" case the positive impact is very slight (Figures 7 and 8), but the forward-looking dynamic structural model suggests that there is more impact than the dynamic structural model with immediate utility maximization.

(3) The forward-looking dynamic structural model suggests that consumers are more willing to try new brands than suggested by the dynamic structural model with immediate utility maximization (see the baseline for the two models in the "New Product" condition). This is because of the information-gathering motive for trying new brands that is captured by the forward-looking dynamic model.

These results are straightforward. They reveal that the two structural models produce very similar results when applied to a product category like laundry detergent, for which the additional value gained by sampling different brands is not large because consumers already know quite a lot about the products. Thus, in the case of mature product categories, it may well be that a structural model that assumes immediate utility maximization can capture most of what a more complicated "dynamic" structural model captures.

6. Discussion and Conclusions

Using Nielsen liquid detergent data, we estimated two structural models of the brand choice behavior of Bayesian consumers in an environment with uncertainty about brand attributes. We also estimate an

Figure 5 Surf Reduces Ad Variability by 50%

(I. Dynamic/New Product)

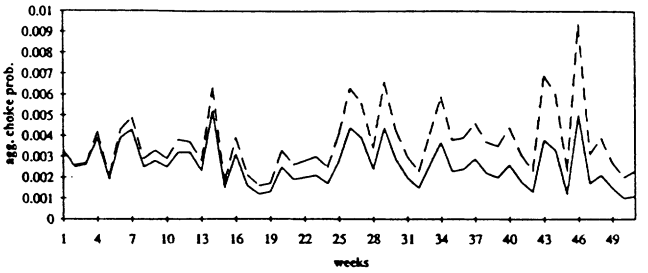


Figure 6 Surf Reduces Ad Variability by 50%

(F. Dynamic/New Product)

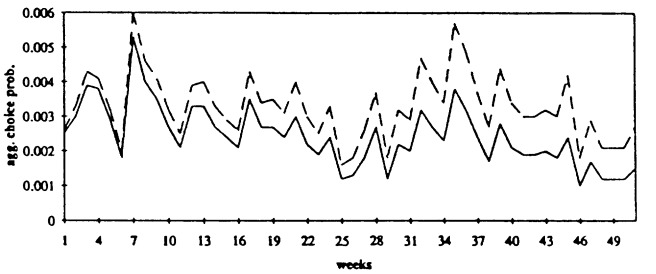


Figure 7 Surf Reduces Ad Variability by 50%

(I. Dynamic/Mature Product)

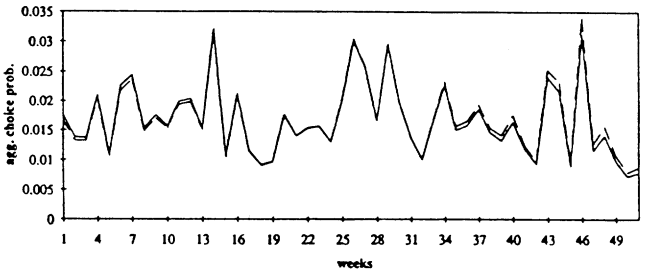
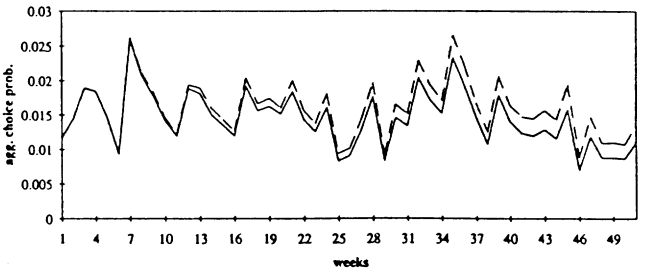


Figure 8 Surf Reduces Ad Variability by 50%

(F. Dynamic/Mature Product)



approximation to the reduced form decision rule that emerges from these models. We refer to the approximation to the reduced form decision rule as the GL model because of its similarity to the Guadagni and Little model (1983).

The main contribution of the paper is to introduce the structural approach to econometric modeling, which has gained wide acceptance in economics during the past decade, into the marketing literature. Within the structural econometrics paradigm, the paper makes two important contributions: (1) it is the first application of the Keane and Wolpin (1994) approximate solution method for DP problems; and (2) it is the first application of structural modelling in a case in which serially correlated error processes must be integrated over in order to solve the consumers' problem. (Such serial correlation arises naturally in Bayesian learning models.)

The proposed structural models incorporate purchase feedback, cumulative advertising effects and consumer expectations about brand attributes as determinants of brand choice. Our structural models provide a theoretical explanation for behavioral variation across households and within households over time. Specifically, we model both usage experience and advertising as sources of information regarding uncertain brand attributes. Over time, households have different experiences when they try brands, and they also may receive different advertising signals. Although consumers have the same priors about brands, their perception errors about mean brand attribute levels diverge over time as they receive different signals. Hence, unobserved heterogeneity of consumers arises endogenously over time. We view this as a great theoretical strength of our models because a behavioral explanation of heterogeneity is provided rather than simply assuming a parametric heterogeneity distribution a priori.

Because heterogeneity arises endogenously over time, our structural models provide a theoretical explanation for brand loyalty formation. Brand loyalty occurs because of the low riskiness of familiar brands. Given a good fit between attributes of a familiar brand and individual tastes, a risk-averse consumer will, ceteris par-

ibus, tend to stay with the familiar brands rather than choosing uncertain alternatives. Because different consumers may have different purchase and advertising exposure histories, they may have different levels of perceived brand risks and perceived brand attribute levels. Indeed, that is why different segments of consumers buy different subsets of brands. Thus, brand loyalty is formed over time via positive use experience and past exposure to low variance (precise) advertising messages.

We find that both structural models fit the data better, both in sample and out of sample, than the GL model. Thus, a key finding of this paper is that a functional form for the dependence of current choices on past experience that was derived explicitly from a theory of Bayesian learning (see Equations 13 and 16) can actually fit detergent choice data as well or better than the flexible exponential smoothing functional form that was assumed by Guadagni and Little (1983) and that has been almost universally applied in the literature on choice dynamics. A functional form such as ours was derived by Roberts and Urban (1988) as well, but it has never been previously applied to scanner panel data. An important avenue for future research is to experiment with this functional form for state dependence in other choice contexts. It is important to note that our structural models represent the first cases in the choice modelling literature in which both the functional forms of state dependence and serial correlation (i.e., heterogeneity) were derived from theory, rather than being specified a priori.

Our structural modelling approach is also useful in analyzing the impact of marketing mix activities on brand choice. Our GL model, which makes the ad hoc assumption that past advertising exposure affects current purchase according to a flexible exponential smoothing functional form, produces estimates that indicate no significant effect of advertising on choice. However, our structural models impose a functional form for the effect of advertising that is consistent with Bayesian learning theory. The structural model estimates indicate that precise (low variance) advertising sustained over a long period creates brand loyalty. In this context, one managerial implication of this work is that signals sent by a firm (in our models, the advertising message) may be imprecise/noisy (i.e., high variance) when a firm's positioning is inconsistent. This inconsistency may have two dimensions: (1) if a firm does not deliver what it promises in its advertising, the signal will be noisy, which will increase overall perceived variance; (2) inconsistency among a firm's messages (reliable family car image versus high status sports car image) also makes signals sent by firms noisy.

Comparing the two structural models, we found that the "forward-looking" dynamic model produced a significantly better in-sample fit than the "immediate utility maximization" dynamic model, but the difference was quantitatively small. In terms of predictive validity (out-of-sample fit), the "immediate utility maximization" dynamic model in which consumers maximize only current period utility performed as well as the "forward-looking" dynamic model. This may be so because there is not much uncertainty about attributes of liquid laundry detergent in spite of new brand introductions, as reflected in our low parameter estimate for initial variance. Consequently, although consumers may be forward-looking, the prevailing uncertainty may not be high enough to motivate consumers to sample different brands frequently in order to learn about them. The "forward-looking" dynamic model will be most useful in product categories where there is a substantial amount of uncertainty and, hence, the expected benefit associated with sampling different brands in order to learn about them is high.

There are many possible extensions to our work. These include allowing for price and advertising frequency as signals of product quality. Also, consumer inventory behavior can be integrated into the current modelling framework, and our modelling of price expectations formation can be enriched.[17]

[17] We would like to thank Jordan Louviere, John Roberts, Peter Rossi, Russell Winer, the Area Editor and two anonymous reviewers for their many suggestions and constructive comments. We also thank the Minnesota Supercomputer Institute and the University of Alberta for their support of this research.

References

Amemiya, Takeshi (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

Bawa, Kapil (1991), "Modeling Inertia and Variety-Seeking Tendencies in Brand Choice Behavior," *Marketing Science*, 3, Summer, 263–278.

Bellman, Richard (1957), *Dynamic Programming*, Princeton, NJ: Princeton University Press.

Carpenter, Gregory S. and Donald R. Lehmann (1985), "A Model of Marketing Mix, Brand Switching and Competition," *Journal of Marketing Research*, 22, August, 318–329.

DeGroot, Morris H. (1970), *Optimal Statistical Decisions*, New York: McGraw-Hill.

Eckstein, Zvi, Dan Horsky and Yoel Raban (1988), "An Empirical Dynamic Model of Brand Choice," Working Paper No. 88, University of Rochester.

Elrod, Terry (1988), "Choice Map: Inferring a Product Market Map from Observed Choice Behavior," *Marketing Science*, 7, Winter, 21–40.

—— and Michael P. Keane (1995), "A Factor-Analytic Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32, February, 1–16.

—— and Russell S. Winer (1991), "An Empirical Comparison of Mapping Methods Based on Panel Data," Working Paper, University of Alberta.

Erdem, Tülin (1993), *Dynamic Brand Choice and Product Market Signalling under Imperfect Information*, Ph.D. thesis manuscript, University of Alberta.

Fishbein, Martin (1967), "Attitude and the Prediction of Behavior," in Martin Fishbein (Ed.), *Readings in Attitude Theory and Measurement*, New York: John Wiley and Sons.

Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, Summer, 203–238.

Hanneman, M. (1984), "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541–561.

Heckman, James B. (1981), "Heterogeneity and State Dependence," in S. Rosen (Ed.), *Studies in Labor Markets*, Chicago, IL: University of Chicago Press, 91–139.

Hoffman, P. J. (1960), "The Paramorphic Representation of Clinical Judgement," *Psychological Bulletin*, 47, 116–131.

Howard, John A. and Jadish N. Sheth (1969), *The Theory of Buyer Behavior*, New York: John Wiley & Sons Inc.

Jeuland, Abel P. (1978), "Brand Preferences over Time: A Partially Deterministic Operationalization of the Notion of Variety-Seeking," in Subhash Jain (Ed.), *AMA Research Frontiers in Marketing: Dialogues and Directions*, Educators' Proceedings, Series No. 43, Chicago, IL: AMA, 33–37.

Kahn, Barbara E., Manohar U. Kalwani and Donald G. Morrison (1986), "Measuring Variety-Seeking and Reinforcement Behaviors Using Panel Data," *Journal of Marketing Research*, 23, May, 89–100.

Keane, Michael P. (1993), "Simulation Estimation for Panel Data Models with Limited Dependent Variables," in G. S. Maddala, C. R. Rao and H. D. Vinod (Eds.), *Handbook of Statistics*, New York: Elsevier Science Publishers.

—— (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 1, 95–116.

—— and Kenneth I. Wolpin (1994), "Solution and Estimation of Dynamic Programming Models by Simulation," *Review of Economics and Statistics*, 76, 4 (November), 648–672.

Lancaster, Kevin (1966), "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 132–157.

Lerman, Steven and Charles Manski (1987) "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric Applications*, C. Manski and D. McFadden (Eds.), Cambridge: MIT Press.

Lucas, R. E., Jr. (1976), *Econometric Policy Evaluation: A Critique*, in K. Meltzer and A. H. Meltzer (Eds.), *The Phillips Curve and Labor Markets*, Supplement to the *Journal of Monetary Economics*.

McFadden, Daniel (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

Meyer, Robert J. and Arvind Sathi (1985), "A Multiattribute Model of Consumer Choice during Product Learning," *Marketing Science*, 4, Winter, 41–61.

Pakes, Ariel (1987), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.

—— and David Pollard (1989), "Simulation and the Asymptotics of Optimization Estimates," *Econometrica*, 57, 1027–1058.

Papatla, Purushottam and Lakshman Krishnamurthi (1992), "A Probit Model of Choice Dynamics," *Marketing Science*, 11, 2, 189–206.

Pedrick, James H. and Fred S. Zufryden (1991), "Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using Single-Source Data," *Marketing Science*, 10, 2 (Spring), 111–130.

Phillips, L. D. and W. Edwards (1966), "Conservatism in a Simple Probability Inference Task," *Journal of Experimental Psychology*, 72, 346–354.

Roberts, John H. and Glen Urban (1988), "Modelling Multiattribute Utility, Risk and Belief Dynamics for New Consumer Durable Brand Choice," *Management Science*, 34, February, 167–185.

Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner," *Econometrica*, 55, 993–1033.

—— (1991), "Estimation of Dynamic Structural Models: Problems and Prospects. Part I: Discrete Decision Processes," Working Paper, SSRI, Madison, WI: University of Wisconsin.

Tellis, Gerard J. (1988), "Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Choice," *Journal of Marketing Research*, 25, May, 134–144.

Wolpin, Kenneth I. (1987), "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, 55, 801–818.

This paper was received September 14, 1993, *and has been with the authors* 8 *months for* 2 *revisions; processed by Robert Meyer, Area Editor.*