Modeling Heterogeneity and State Dependence in Consumer Choice Behavior

Michael P. Keane
Department of Economics, University of Minnesota, Minneapolis, MN 55455

In frequently-purchased-consumer-goods markets, consumer brand choices exhibit substantial persistence across purchase occasions. In this article, I estimate a choice model that admits of both heterogeneity in preferences and true state dependence as sources of this persistence, using Nielsen scanner data on ketchup. I find evidence for true state dependence in the choice process, even after controlling for a rich heterogeneity structure. Simulation of the model indicates that the long-term effect of a promotion-induced purchase on future purchase probabilities is positive but small.

KEY WORDS: Brand choice; Discrete choice model; Method of simulated moments; Multinomial probit; Scanner data.

In this article I estimate statistical models of consumer choice behavior using Nielsen scanner panel data on household ketchup purchases. The goal is to determine whether *heterogeneity* or *state dependence* or both are important sources of the observed temporal persistence of consumer brand choices and to gauge the relative importance of each. Most recent statistical models of choice behavior have allowed for either heterogeneity or state dependence, but rarely have both sources of persistence in choices been included simultaneously. My tentative conclusion is that both heterogeneity and state dependence appear to be significant features of the data. But before proceeding it is important to carefully define both terms.

A basic fact about panel data on consumer-goods purchases is that brand choices of individual consumers exhibit persistence over time. More precisely, we observe the following: The probability that a consumer will be observed to buy brand A at time t, given that he/she was observed to buy brand A at time $t - 1$, exceeds the probability that a consumer will be observed to purchase brand A at time t, given that he/she was observed to purchase brand B at time $t - 1$. The word "observed" in this statement is critical because it indicates that we are speaking from the perspective of the analyst. For this reason, this fact is merely a statement about how the data look from the perspective of an analyst—it is not a statement about consumer behavior.

The fact that brand choices of consumers exhibit persistence is consistent with two diametrically opposed patterns of consumer behavior. First, it may be that purchase of a particular brand at time $t - 1$ actually makes the consumer more likely to purchase that brand again at t. There are a myriad of plausible explanations for such a causal link between current and past behavior. For example, there may be habit persistence (i.e., use of a brand causes one to acquire a taste for that brand), learning combined with risk aversion (i.e., use of a brand gives one knowledge about its attributes, making it a safe choice on a subsequent purchase occasion), and so forth. Regardless of the underlying behavioral explanation, I shall refer to such causal links between past and present purchase behavior as state dependence. The phrases *purchase carryover* and *purchase-event feedback* are often used in the marketing literature to describe the same general concept.

The second pattern of behavior with which the observed persistence in consumer brand choices is consistent is one in which past purchase behavior has no causal connection with present purchase behavior. Rather, consumers simply have different preferences over brands for exogenous reasons that are unrelated to the consumers' past purchase histories. In this case, a consumer who was observed to buy brand A at time $t-1$ is more likely to have preferences such that he/she generally prefers brand A than is a consumer who bought brand B at time $t - 1$. For this reason alone, a consumer who bought A at $t - 1$ is more likely to buy A at t than is a consumer who bought B at $t - 1$. I shall refer to such differences in exogenously given preferences of consumers as heterogeneity.

Distinguishing between heterogeneity and state-dependence-based explanations for the observed persistence in consumer brand choices is of fundamental importance in marketing. Consider, for example, the decision to put a particular brand on price promotion. If the true model of consumer behavior is that there is heterogeneity and no state dependence, such a promotion will increase sales only while it is in effect. If state dependence is present, however, some of the consumers who bought the brand on promotion will be persuaded to stay with the brand after the price returns to normal. A cost/benefit analysis of the promotion will depend critically on the assumed forms of heterogeneity and state dependence in the population.

Having defined heterogeneity and state dependence, my next task is to discuss how these diametrically opposed explanations for the observed persistence in consumer choice behavior may be distinguished using scanner data on consumer purchases. As is well known (see Heckman 1981a), if heterogeneity is present in the true model and one ignores it, estimating a model that only allows for state dependence, one will tend to overestimate the degree of state dependence. This is referred to as a finding of *spurious state dependence*. Equally true, but less often noted, is the fact that if state dependence is present in the true model and one ignores it, estimating a model with only heterogeneity,

Journal of Business & Economic Statistics
July 1997, Vol. 15, No. 3

then the degree of heterogeneity in the population will tend to be exaggerated.

Clearly then, we want to estimate a model that allows for both heterogeneity and state-dependence explanations of the observed persistence in brand choices so as to disentangle the contribution of each. Unfortunately, there is no nonparametric method to sort out the relative importance of heterogeneity and state dependence in generating persistence. Inferences about the relative and absolute importance of heterogeneity and state dependence are conditional on functional-form assumptions. And a researcher has an immense range of options in the specification of these forms. I next turn to a discussion of these options.

In a random-utility model, state dependence can be accounted for by allowing past purchases to have an impact on current-period utility evaluations (Guadagni and Little 1983) or letting attributes of previously purchased brands have an impact on current utility evaluations (Erdem 1996). Heterogeneity can be accounted for by letting certain parameters of the utility function differ across consumers. One may let parameters differ according to observed attributes of consumers, unobserved attributes of consumers (Elrod 1988; Jones and Landwehr 1988; Steckel and Vanhonacker 1988) or both (McCulloch and Rossi 1996). Furthermore, one can even allow heterogeneity in the way past purchases affect current decisions (Erdem 1996). In terms of functional forms for dependence of current utility evaluations on past choices, one can use a dummy for lagged purchase (Jones and Landwehr 1988), an exponentially smoothed weighted average of past purchases (Guadagni and Little 1983), various Bayesian updating formulas (Fader 1993; Erdem and Keane 1996), or one of a myriad of other possible forms. In terms of functional forms for heterogeneity distributions, one can choose a normal distribution (Elrod 1988; Elrod and Keane 1995), a gamma distribution (Bass, Jeuland, and Wright 1976), a discrete mass point distribution (Kamakura and Russell 1989), or one of a myriad of other distributions. One can also specify that a consumer's position in the heterogeneity distribution is fixed over time or allow it to evolve exogenously over time (Allenby and Lenk 1994).

This list of options and references is not meant to be exhaustive but merely to give an idea of the range of options available in specifying functional forms for heterogeneity and state dependence. Given the wide range of model-specification options, my strategy for proceeding is as follows: I will specify the particular functional form for state dependence that is dominant in the marketing literature on consumer brand choice, the exponential smoothing form used by Guadagni and Little (1983). I will then estimate models with successively more flexible functional forms for heterogeneity and see if the parameters capturing state dependence of the Guadagni–Little (GL) form are significant in each of these models. If the GL parameters remain significant even after a complex pattern of heterogeneity is controlled for, I will take it as evidence that state dependence is an important aspect of consumer choice behavior. On the other hand, if the inclusion of reasonably flexible functional forms for heterogeneity renders the GL parameters insignificant, I will take it as evidence that almost all observed persistence in consumer choice behavior is due to heterogeneity.

The conventional wisdom in marketing is that choice behavior is *zero order*, meaning that there is no significant causal connection between past purchases and current utility evaluations [see Bass (1993) for a statement of this position]. In fact, in the Nielsen scanner data on ketchup purchases that I have examined, this turns out not to be the case. The GL parameters remain highly significant even after inclusion of an extremely flexible specification of heterogeneity. I take this as evidence that state dependence is present in consumer choice behavior.

But, in an important sense, my results are quite consistent with the conventional wisdom. According to my preferred model specification, a price promotion for a brand of ketchup that is large enough to generate a 313% immediate sales increase only increases long-run sales after price returns to normal by about 12%. Thus, the effect of current prices on current utility evaluations is much stronger than the effect of past purchases. As a result, when one looks at scanner data, the price effects will tend to swamp any purchase carryover effects, making zero-order choice behavior a good first approximation. Only highly efficient econometric methods, like those employed here, allow one to tease out the state dependence in the choice process.

The article is organized as follows: The model, estimation technique, and data are discussed in Sections 1, 2, and 3, respectively. Results are presented in Sections 4, 5, and 6. In Section 7, I use some of the estimated models to simulate consumers' dynamic responses to price promotions, showing how responses differ in models with and without state dependence. Section 8 concludes.

1. THE MODEL

1.1 Mathematical Structure

Assume that in each time period (which for our purposes will be the purchase occasion) consumers choose from among a set of J brands. Let U_{ijt} denote the utility to consumer i of purchasing brand j at time t. This utility will be allowed to depend on observed and unobserved characteristics of brands, observed and unobserved characteristics of the consumer (including his/her observed brand-choice history), and interactions among consumer and brand characteristics. Specifically, let

$$U_{ijt} = \mathbf{X}_{it}\boldsymbol{\beta}_j + \mathbf{A}_{jt}(\phi_0 + \mathbf{X}_{it}\phi_1 + \nu_i) + \mathrm{GL}(H_{ijt}, \alpha)\lambda + \mathbf{A}^U_{jt}\boldsymbol{\Omega}_{ijt}, \qquad j = 1, J. \quad (1)$$

Here, $\mathbf{X}_{it}$ is a vector of characteristics of consumer i at time t, and $\boldsymbol{\beta}_j$ is a corresponding vector of coefficients capturing how these consumer attributes affect the consumers' evaluation of the utility from brand j. $\mathbf{A}_{jt}$ is a vector of observed attributes of brand j at time t. It is multiplied by a vector of random coefficients $\phi_0 + \mathbf{X}_{it}\phi_1 + \nu_i$ that captures how consumers with different observed and unobserved characteristics evaluate the utility derived from these attributes.

$\mathbf{A}_{jt}^{U}$ is a vector of unobserved attributes of brand j at time t, and $\boldsymbol{\Omega}_{ijt}$ is a vector of random coefficients that captures how consumers evaluate the utility derived from the unobserved attributes. $\mathrm{GL}(H_{ijt}, \alpha)$ is an exponentially smoothed weighted average of past purchases of brand j by person i used by Guadagni and Little (1983), H_{ijt} is consumer i's purchase history for brand j prior to time t, and α is the exponential smoothing parameter; λ is the coefficient mapping GL into the evaluation of utility.

Many alternative error structures may be obtained by specifying the functional form of $\mathbf{A}_{jt}^{U}\boldsymbol{\Omega}_{ijt}$. Following Elrod and Keane (1995), I interpret the error terms as arising from unobserved attributes of brands for which consumers have heterogeneous preferences. These attributes are assumed to be of two types—type I attributes, for which consumers have utility weights that are fixed over time, and type II attributes, for which consumers have utility weights that vary over time. For example, a type I attribute might be freshness, on which consumers place a time-invariant weight, whereas a type II attribute might be a kind of flavor, which consumers may desire on some days but not on others.

Moreover, attributes may be common across brands or unique to brands. A common attribute is something that can be measured for all brands, such as meat content, freshness, or glamour. A unique attribute is inherently undefinable (e.g., the unique attribute of Budweiser is whatever gives Budweiser its "Budweiserness," independent of its position along the common-attribute dimensions).

This gives the following very general structure:

$$\mathbf{A}_{jt}^{U}\boldsymbol{\Omega}_{ijt} = \mathbf{L}_j\mathbf{W}_i + \sqrt{\kappa_j}\Gamma_{ij} + \mathbf{P}_j\boldsymbol{\xi}_{it} + \sqrt{\tau_j}\varepsilon_{ijt}, \qquad j = 1, J. \quad (2)$$

Here W_i is the vector of utility weights attached by consumer i to the type I common attributes; L_j is the vector of factor loadings of brand j on these type I common attributes; Γ_{ij} is the utility weight attached by consumer i to the type I unique attribute of brand j, $\sqrt{\kappa_j}$ is the level of type I unique attribute possessed by brand j; $\boldsymbol{\xi}_{it}$ is the time t vector of utility weights attached by consumer i to the type II common attributes of brands; $\mathbf{P}_j$ is the vector of factor loadings of brand j on these type II common attributes; ε_{ijt} is the time t utility weight attached by consumer i to the type II unique attribute of brand j; $\sqrt{\tau_j}$ is the level of type I unique attribute possessed by brand j.

It is important to note that the error structure for all random-utility-based brand-choice models that have been considered in the literature may be obtained by placing suitable restrictions on (2). This is because any covariance matrix may be factor analyzed as in (2). For example, if $\mathbf{L}_j = 0$, $\sqrt{\kappa_j} = 0$, and $\mathbf{P}_j = 0$ for all j, $\sqrt{\tau_j} = 1$ for all j, and ε_{ijt} is iid extreme value, we obtain the logit model. If $\sqrt{\kappa_j} = 0$, and $\mathbf{P}_j = 0$ for all j, $\sqrt{\tau_j} = 1$ for all j, ε_{ijt} is iid extreme value, and $\mathbf{W}_i$ is iid normal, we obtain the heterogeneous logit model of Elrod (1988). Alternatively, if $\mathbf{L}_j\mathbf{W}_i$ is assumed to have a discrete mass point distribution, we obtain the heterogeneous logit model of Kamakura and Russell (1989), but if $\mathbf{W}_i$ is assumed to have a discrete mass point distribution and the elements of the $\mathbf{L}_j$ are estimated, we obtain the heterogeneous logit model of Chintagunta (1994). And, if $\mathbf{L}_j\mathbf{W}_i$ is allowed to be an individual fixed effect, we have the fixed-effects logit model of Jones and Landwehr (1988). If $\mathbf{L}_j = 0$ and $\mathbf{P}_j = 0$ for all j, $\sqrt{\kappa_j} = \sqrt{\kappa}$ for all j, $\sqrt{\tau_j} = 1$ for all j, Γ_{ij} is gamma, and ε_{ijt} is iid extreme value, we obtain the Dirichlet-multinomial model (Bass et al. 1976). If $\mathbf{L}_j = 0$, $\sqrt{\kappa_j} = 0$, and $\mathbf{P}_j = 0$ for all j, $\sqrt{\tau_j} = 1$ for all j, and ε_{ijt} is iid normal, we obtain the iid probit model. Letting $\mathbf{P}_j$ be nonzero and $\boldsymbol{\xi}_{it}$ be iid normal gives a cross-section probit model with correlated errors, and further letting $\mathbf{L}_j$ be nonzero and $\mathbf{W}_i$ be iid normal gives a random-effects probit model. The heterogeneous iid probit model of Elrod and Keane (1995) is obtained by letting $\sqrt{\kappa_j} = \sqrt{\kappa}$ for all j and $\sqrt{\tau_j} = 1$ for all j, letting $\mathbf{W}_i$, $\boldsymbol{\Gamma}_{ij}$, and ε_{ijt} all be iid standard normal and letting $\mathbf{P}_j = 0$ for all j. This list may be extended almost indefinitely.

It has been nearly universal in the brand-choice literature to work with permanent/transitory models, meaning that a stark dichotomy is set up between $\mathbf{W}_i$ and Γ_{ij}, which are fixed over time, and $\boldsymbol{\xi}_{it}$ and ε_{ijt}, which are serially independent. This leads to an equicorrelated error structure (i.e., the correlation of the error terms $\mathbf{A}_{jt}^{U}\boldsymbol{\Omega}_{ijt}$ and $\mathbf{A}_{j,t-q}^{U}\boldsymbol{\Omega}_{ij,t-q}$ is the same regardless of the lag length q). This assumption that a consumer's position in the heterogeneity distribution is constant over time (i.e., that it is given at birth and never changes) is highly restrictive, and it has been rejected in all contexts in which it has been tested (see Heckman 1981a; Avery, Hansen, and Hotz 1983; Keane 1993). To my knowledge, the only cases in the brand-choice literature in which equicorrelation has been relaxed are those of Allenby and Lenk (1994) and Fader and Lattin (1993).

In this article I will assume that $\sqrt{\kappa_j} = \sqrt{\kappa}$ for all j and $\sqrt{\tau_j} = 1$ for all j and that $\mathbf{W}_i$, Γ_{ij}, $\boldsymbol{\xi}_{it}$, and ε_{ijt} are all iid standard normal. This gives a multinomial multiperiod probit (MMP) model with a very flexible covariance structure. I will also relax equicorrelation by allowing the time-varying component of the error terms to follow a stationary (autoregressive) AR(1) process. Specifically, define

$$\delta_{ijt} = \mathbf{P}_j\boldsymbol{\xi}_{it} + \sqrt{\tau_j}\varepsilon_{ijt}, \qquad j = 1, J, \quad (3)$$

and let

$$\delta_{ijt} = \rho\delta_{ij,t-1} + \eta_{ijt}, \qquad j = 1, J, \quad (4)$$

where $0 \leq \rho < 1$ and the η_{ijt} are iid normal over time but have the appropriate cross-sectional correlation across brands so as to preserve a stationary covariance structure for δ_{ijt}.

Note that the $\sqrt{\tau_1} = 1$ restriction serves as a scale normalization for utility, which is needed for identification (see Sec. 2.1). Both the $\sqrt{\kappa_j} = \sqrt{\kappa}$ for all j and $\sqrt{\tau_j} = 1$ for $j \geq 2$ restrictions were imposed because differences across brands in the levels of unique attributes were not found to be significant in the estimation and including these additional parameters led to a significant increase in computational burden. I have estimated a model that includes only one column in both the $\mathbf{L}$ and $\mathbf{P}$ vectors, meaning that there is one unobserved type I common attribute and one unobserved type II common attribute. Additional columns of $\mathbf{L}$ and $\mathbf{P}$ were not found to be significant.

Finally, I assume that the heterogeneous part of the observed attribute coefficient $\phi_0 + \mathbf{X}_{it}\phi_1 + \nu_i$ is 0 for all observed attributes except price. For the price coefficient I will include household income and household size in $\mathbf{X}_{it}$ and assume $\nu_i \sim N(0, \sigma_\nu^2)$ with the ν_i independent across alternatives. [Note that Kamakura and Russell (1989) and Allenby and Lenk (1994) allowed for correlation between a random price coefficient and $\mathbf{L}_j\mathbf{W}_i$. Given the additional complexities included in the present model, such correlations are omitted to conserve on parameters. Nevertheless, the presence of the $\mathbf{X}_{it}\boldsymbol{\beta}_j$ and $\mathbf{X}_{it}\phi_1$ terms allows household price coefficients and preferences for brands to be correlated due to observable household characteristics.] The details of the specification of the observed attribute vector $\mathbf{A}$ will be described in Section 3.

1.2 Behavioral Interpretation

Having laid out the complete statistical model, it is useful to give behavioral interpretations of the parameters. Consider first the covariance structure. If $\mathbf{L}_j$ is very large, then those consumers with a large positive value of $\mathbf{W}_i$ will appear loyal to brand j (i.e., those consumers who place a large weight on the type I common attribute will tend to buy brand j because it has a high level of that attribute). If $\sqrt{\kappa}$ is large, then each consumer will appear loyal to a particular brand for no apparent reason. There is thus a crucial managerial distinction between markets in which the $\sqrt{\kappa_j}$ are large (κ heterogeneity) and markets where the $\mathbf{L}_j$ are large ($\mathbf{L}$ heterogeneity). Both markets will be characterized by brand loyalty and strong persistence in brand-choice behavior. In the case of $\mathbf{L}$ heterogeneity, however, a manager could easily steal consumers away from a competing brand by repositioning his/her brand along the $\mathbf{L}$ vector, but in the case of κ heterogeneity it is not clear how a brand manager could steal loyal consumers from another brand.

The elements of the $\mathbf{L}$ vector influence brand-switching behavior. If $\mathbf{L}_j$ and $\mathbf{L}_k$ are similar, it means that brands j and k have similar type I common-attribute levels, and we would expect to see a segment of consumers who switch frequently between these brands. Conversely, if $\mathbf{L}_j$ and $\mathbf{L}_k$ are very different, it means that j and k have very different levels of the type I common attribute, and we would not expect to see any segment of consumers who switch frequently between these brands.

The elements of the $\mathbf{P}$ vector also influence switching behavior. If $\rho = 0$ so that a consumer gets an independent draw for the $\boldsymbol{\xi}_{it}$ vector on each purchase occasion, consumers will tend to switch between brands that have large elements of $\mathbf{P}$ that are of opposite sign. In an extreme case in which $\mathbf{P}_j \to \infty$ and $\mathbf{P}_k \to -\infty$, we will observe all consumers switching back and forth between j and k, with the purchase probabilities for each approaching 50%.

If ρ is substantially greater than 0 and $\mathbf{P}_j$ is large in magnitude, we will see stretches of purchase occasions when consumers appear loyal to brand j and stretches when they become averse to j. In other words, brand j has a high level of a common attribute for which consumers go through phases such that they like that attribute over some periods of time and dislike it over other periods of time. Alternatively, if $\rho \gg 0$ and all other model parameters (including the $\mathbf{P}$) are small relative to $\sqrt{\tau} = 1$, consumers will tend to go through phases when they appear loyal to all the brands in turn.

It is important to recognize the crucial behavioral distinction between a large ρ and large amounts of $\mathbf{L}$ heterogeneity and κ heterogeneity. In the latter case, consumer behavior (in terms of which brand or brands a particular consumer most prefers) will be static over indefinitely long periods of time. In the large ρ case, however, consumer behavior will only tend to be similar over short periods of time rather than over indefinite periods. In other words, with large ρ there is "inertia" in choice behavior over the short to medium run, but not in the long run.

Altogether, the model specified in Equations (1)–(4) includes seven types of heterogeneity. These are (1) observed heterogeneity in preferences for observed attributes, captured by $\mathbf{A}_{jt}\mathbf{X}_{it}\phi_1$, (2) unobserved heterogeneity in preferences for observed attributes, captured by $\mathbf{A}_{jt}\nu_i$, (3) observed heterogeneity in brand intercepts, captured by $\mathbf{X}_{it}\boldsymbol{\beta}_j$, and (4–7) the four types of unobserved heterogeneity in intercepts captured by the four terms on the right side of Equation (2). These are heterogeneous preferences for type I and II common attributes and type I and II unique attributes.

State dependence is captured by the term $\mathrm{GL}(H_{ijt}, \alpha)\lambda$. If $\lambda = 0$, then choice is a zero-order process, with all observed persistence due to heterogeneity. But, if $\lambda > 0$, purchase of a brand at t causes the consumer to assign a higher utility to purchase of that brand at $t+1$. Thus, if and only if $\lambda > 0$ will there be purchase carryover effects such that a promotion that increases sales at t can lead to increased sales at $t+1$ after the promotion is removed. In contrast, recent work by Allenby and Lenk (1994) allowed for a complex heterogeneity structure including AR(1) errors of the type considered here but did not allow for state dependence. They estimated $\rho \gg 0$ and referred to this phenomenon as "purchase carry-over." But this is quite different from my use of the term. In their model, consumers tend to purchase the same brand on successive occasions because of short-run inertia in preferences, not because the time t purchase is causally related to time $t+1$ utility evaluations.

2. ESTIMATION

2.1 Identification

Because choices among alternative brands only depend on utility differences, to achieve identification assume that all the U_{ijt}, $\mathbf{X}_{it}$, $\mathbf{A}_{jt}$, GL_{ijt}, and $\mathbf{A}_{jt}^U$ in Equation (1) are measured relative to brand J so that $U_{iJt} = 0$. Note, then, that the positions of alternatives along the $\mathbf{L}$ and $\mathbf{P}$ vectors are measured relative to alternative J and that $\mathbf{L}_J = \mathbf{P}_J = 0$.

After this differencing procedure, the error term for alternative j is $\mathbf{A}_{jt}^U\boldsymbol{\Omega}_{ijt} - \mathbf{A}_{Jt}^U\boldsymbol{\Omega}_{iJt}$. The total variance of this composite error term is $2\kappa + \mathbf{L}_j^2 + \mathbf{A}_{jt}^2\sigma_\nu + (2\tau_j + \mathbf{P}_j^2)/$

$(1-\rho^2)$. Because it is possible to divide $\boldsymbol{\beta}_j$ for all j, ϕ_0, ϕ_1, ν_i, λ, and the error terms for all j by a constant without changing the rankings of alternatives, a scale normalization for utility is also needed. This is achieved by normalizing the error variance for alternative 1 to 1 at $t=1$. Then, the total error variance for alternative j is $[2\kappa+\mathbf{L}_j^2+\mathbf{A}_{jt}^2\sigma_\nu^2+(2\tau_j+\mathbf{P}_j^2)/(1-\rho^2)]/[2\kappa+\mathbf{L}_1^2+\mathbf{A}_{11}^2\sigma_\nu^2+(2\tau_1+\mathbf{P}_1^2)/(1-\rho^2)]$. At this point, identification is not yet achieved because both numerator and denominator of this expression may be multiplied by the same constant without affecting utilities. The additional normalization I use is $\tau_j=1$ for all j. This is with loss of generality because all that would be necessary is $\tau_1=1$. As I noted in Section 1, however, differences in the τ_j across alternatives were not found to be significant.

With this normalization, the fraction of the total error variance for alternative j due to κ heterogeneity (i.e., population heterogeneity in preferences for the unique type I attribute of alternative 1) is $2\kappa/[2\kappa+\mathbf{L}_j^2+\mathbf{A}_{jt}^2\sigma_\nu^2+(2\tau_j+\mathbf{P}_j^2)/(1-\rho^2)]$. Fractions of error variance due to other sources are obtained similarly. Finally, note that with the τ_j restriction, the alternative 1 error-variance normalization actually becomes superfluous. It leads to no loss in generality, however, and has the advantage that the scales of the $\boldsymbol{\beta}$, ϕ_0, and λ coefficients will remain reasonably comparable across specifications with different heterogeneity specifications.

2.2 Estimation

Estimation of an MMP model with an error structure as complex as that described in Section 1 requires the use of simulation estimation techniques that I describe in this section. Assume that we have data on N consumers who choose from among a set of J brands on each of T_i purchase occasions. Choice j is made at time t if the following $J-1$ constraints are satisfied: $U_{ijt}>U_{ikt}$ for all $k\neq j$. Write the model in compact notation as $U_{ijt}=\mathbf{X}_{ijt}^*\boldsymbol{\beta}_j^*+\varepsilon_{ijt}^*$, where $\mathbf{X}_{ijt}^*\boldsymbol{\beta}_j^*=\mathbf{X}_{it}\boldsymbol{\beta}_j+\mathbf{A}_{jt}(\phi_0+\mathbf{X}_{it}\phi_1)+\mathrm{GL}(H_{ijt},\alpha)\lambda$ and $\varepsilon_{ijt}^*=\mathbf{A}_{jt}\nu_i+\mathbf{A}_{jt}^U\boldsymbol{\Omega}_{ijt}$. Further define $\varepsilon_i^*=(\varepsilon_{i11}^*,\ldots,\varepsilon_{i,J-1,1}^*,\ldots,\varepsilon_{i1T}^*,\ldots,\varepsilon_{i,J-1,T}^*)'\sim\mathrm{IIDN}(0,\Sigma^*)$. Let $\theta=(\boldsymbol{\beta},\phi_0,\phi_1,\alpha,\lambda,\mathbf{L},\kappa,\Gamma,\mathbf{P},\sigma_\nu^2)$ denote the complete vector of model parameters. To have a compact notation for the sequence of choices observed for person i, define $d_{it}=(d_{i1t},\ldots,d_{iJt})$, and $d_i=(d_{i1},\ldots,d_{iT})$. Also define j_{it} as the index j of the choice observed for person i in period t. Denoting by $P(d_i)$ the probability that i chooses the sequence d_i, we have

$$\begin{aligned}
&P(d_i|\theta,\mathbf{X},\mathbf{A})\\
&\quad= P(U_{i,j_{it},t}>U_{ikt}\forall\ k\neq j_{it},t=1,\ldots,T)\\
&\quad= P(\varepsilon_{i,j_{it},t}^*-\varepsilon_{ikt}^*>\mathbf{X}_{ikt}^*\boldsymbol{\beta}_j^*-\mathbf{X}_{i,j_{it},t}^*\boldsymbol{\beta}_{j_{it}}^*\\
&\qquad\quad \forall\ k\neq j_{it},t=1,\ldots,T).
\end{aligned}$$

If the ε_{ijt}^* are iid over time, then this is the product of T integrals of dimension $J-1$. If the ε_{ijt}^* are serially correlated, however, this is in general a $T\cdot(J-1)$ variate integral. As T and/or J grow, the repeated evaluation of such integrals that is necessary for maximum likelihood estimation of the model rapidly becomes infeasible. Much of the earlier work on the MMP model sought to avoid this problem by imposing low-order factor structures on Σ^*. For example, if a random-effects structure is imposed, the order of integration is reduced to $(J-1)\cdot 2$.

Unfortunately, given the AR(1) error component in the model of Section 1, the order of integration necessary to evaluate choice probabilities in the model will be the full $T\cdot(J-1)$. In the empirical example to be presented in Sections 3–7, $T=30$ and $J=7$, giving an order of integration of 180. For this reason I adopt a simulation-based approach to inference.

Most classical simulation-based approaches to inference in the MMP model rely on Monte Carlo simulation of the choice sequence probabilities $P(d_i)$ and substitution of these simulated probabilities into likelihood functions or moment conditions. A critical decision is which Monte Carlo method to use. In an extensive study of alternative methods for simulation of multinomial orthant probabilities, Hajivassiliou, McFadden, and Ruud (1994) concluded that the GHK probability simulator, due to Geweke (1991), Hajivassiliou et al. (1994), and Keane (1990, 1994), is the most accurate of all methods considered. For this reason, I adopt the GHK method. Denote the GHK simulator of choice sequence d_i by $\hat{P}_{\mathrm{GHK}}(d_i|\theta,\mathbf{X},\mathbf{A})$. For a description of the construction of the GHK simulator in the panel-data probit case, see Geweke, Keane, and Runkle (in press).

The next decision is the choice of estimator. I adopt the method of simulated moments (MSM). As discussed by Keane (1990), direct application of the MSM estimator developed by McFadden (1989) to the MMP model is not feasible for $T>2$. Keane (1990) proposed the computationally feasible alternative of factoring the sequence probabilities into transition probabilities and forming the alternative MSM estimator obtained by solving the moment conditions:

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}\mathbf{W}_{ijt}[d_{ijt}-\hat{P}_{\mathrm{GHK}}\\
&\qquad\times(d_{ijt}|d_{i1},\ldots,d_{i,t-1},\theta_{\mathrm{MSM}},\mathbf{X},\mathbf{A},H_{it})]=0,
\end{aligned}$$

where $\mathbf{W}_{ijt}$ is an orthogonal weight, and the transition probabilities are simulated by ratios of GHK simulators:

$$\hat{P}_{\mathrm{GHK}}(d_{ijt}|d_{i1},\ldots,d_{i,t-1})\equiv\frac{\hat{P}_{\mathrm{GHK}}(d_{i1},\ldots,d_{i,t-1},d_{ijt})}{\hat{P}_{\mathrm{GHK}}(d_{i1},\ldots,d_{i,t-1})}.$$

Due to denominator bias, this gives a biased simulator of the transition probability. Keane (1994) showed that this MSM estimator is consistent and asymptotically normal if $M/\sqrt{N}\to\infty$ as $N\to\infty$, however, and he found in a Monte Carlo study that it has good small-sample properties.

Although the high-order integration problem in MMP models is well known, an additional important problem is proliferation of covariance matrix parameters. In an unrestricted MMP model, there are $T(J-1)J/2-1$ free covariance matrix parameters (i.e., covariances of all error terms across all time periods). In the present case, $T=30$ and $J=7$, so this gives 631. Estimation of 631 free parameters via a nonlinear estimation routine in which each objective function evaluation requires high-dimensional integration is completely infeasible. Furthermore, given the limited information content of discrete choice data, it is clear that data on many thousands of individual purchase histories would be needed before one could hope to uncover the parameters of a completely unrestricted structure.

The statistical model of Section 1 circumvents this proliferation of parameters problem while nevertheless allowing for a flexible correlation structure in the errors. With the restrictions that $\mathbf{L}$ and $\mathbf{P}$ have one column each, that $\sqrt{\kappa_j} = \sqrt{\kappa}$ and $\sqrt{\tau_j} = 1$ for all j, and that $\sigma_\nu = 0$ for all $\mathbf{A}$ except price, the number of parameters characterizing the covariance matrix is only 15.

A final problem that must be confronted in estimation of a limited dependent-variable model with state dependence is that of initial conditions. Let $t = 1$ denote the first purchase occasion for which we observe the consumer. The consumer first entered the market for the product under consideration, however, at $t = -B$, where $B+1$ is the number of purchase occasions prior to our first observation. Because we do not observe the history of the consumer's choice process from $t = -B$ to $t = 0$ and because we do not observe the latent utilities, we cannot construct either the initial value of the loyalty variable $\mathrm{GL}(H_{ij1}, \alpha)$ or of the error term δ_{ij0}.

Given this initial-conditions problem, estimation that is consistent in sample size N for fixed T requires integration over the joint unconditional distribution of $\mathrm{GL}(H_{ij1}, \alpha)$ and δ_{ij0}. As described by Heckman (1981b), this is a problem of awesome computational and economic complexity. One must integrate over the entire past history of the exogenous variables and of the consumer's decisions, starting from time $t = -B$. This requires one to model the complete structure of the market, including how prices are set, promotions are determined, brands are born and die, and so forth.

Estimation that is consistent as N and T both grow large may be achieved in two ways. First, one may treat the unobserved $\mathrm{GL}(H_{ij1}, \alpha)$ as individual fixed effects. This generates N new parameters. Estimation of such a large set of free parameters in a nonlinear estimation routine in which each objective function evaluation requires high-dimensional integration is completely infeasible. Alternatively, if $\mathrm{GL}(H_{ij1}, \alpha)$ is arbitrarily initialized, estimation remains consistent as both N and T grow large. This is because, as T grows large, results become insensitive to initial conditions.

By force of necessity I adopt the approach of arbitrarily initializing $\mathrm{GL}(H_{ij1}, \alpha)$ to be 0 for all brands. This has the desirable feature of ensuring that the assumed initial conditions for the GL variables are exogenous in the sense that they are uncorrelated with consumers' positions in the heterogeneity distribution. Thus, with this initialization, the lagged purchase variable cannot be significant simply because its initial value is correlated with individuals' unobserved preference parameters. But with $\mathrm{GL}(H_{ij1}, \alpha)$ set to 0, ε_{ij1} contains information both on the consumer's position in the heterogeneity distribution and on $\mathrm{GL}(H_{ij1}, \alpha)$. As a result, consistency will only be achieved as T grows large. I check the sensitivity of the estimates to T by estimating the model both on the full sample and on a subsample of consumers who have relatively long purchase strings.

3. DATA

The data used in the study are the Nielsen scanner panel data on ketchup for Sioux Falls, South Dakota. The sample period is 1987, week 16, to 1988, week 24. During this period there were seven name-brand varieties of ketchup that accounted for the bulk of sales in the market. These are Hunt's (32 oz.), Del Monte (32 oz.), and five sizes of Heinz (40, 64, 14, 28, and 32 oz.). The Heinz 40-ounce size was only introduced in 1987, week 16, whereas the other alternatives had been present in the market earlier. Each of these seven will be considered a choice alternative in this study. Market shares are given in Table 1. The dominant alternative, with a market share of 33.8%, is the Heinz 32-ounce size.

The most difficult task involved in using scanner panel data is the construction of the vector of prices faced by each consumer on each purchase occasion. The basic problem is that one only observes the price paid by the consumer for the brand he/she actually bought. This is broken down separately into a marked price and the value of any coupon that was used. Prices for the other brands must be inferred. The Nielsen data include price files that contain prices for each brand in each store on each day of the sample. Details of the construction of these prices can be found in the documentation that accompanies the Nielsen data. But the basic idea of the algorithm is as follows: (1) Sort through all the data for a particular store on a particular day. If a consumer is found who bought a particular brand, then use the marked price he/she faced as the marked price for that brand in that store on that day. (2) If no one bought a particular brand in a particular store on a particular day, look for purchases in adjacent days to fill in the price. (3) If no one bought a particular brand in a particular store in a particular week, then look for purchases in adjacent weeks to fill in the price. Prices for all brands are measured as price per 32 ounces.

In constructing the price data, it is important to use

Table 1. Characteristics of Alternatives in the Sioux Falls Ketchup Data

Alternative	*Brand*	*Weight (ounces)*	*Mean price per 32 oz.*	*Container type*	*Market share*
1	Hunt's	32	1.10	?*	16.5
2	Del Monte	32	1.10	Glass	6.1
3	Heinz	40	1.35	Plastic	8.2
4	Heinz	64	1.32	Plastic	3.3
5	Heinz	14	1.52	Glass	3.5
6	Heinz	28	1.28	Plastic	28.6
7	Heinz	32	1.12	Glass	33.8

NOTE: Purchases of house brands and generics are excluded from the sample, so market share refers to share among the seven name-brand alternatives. In Sioux Falls the total market share of all house brands and generics combined was approximately 6%.

* For Hunt's (32 oz.) the UPC code did not uniquely identify whether the package was plastic or glass, and I was informed by Nielsen that both plastic and glass containers were available. Thus, the plastic dummy is set to .50.

the marked price for the purchased brand rather than the price net of any coupon redemption. Because coupon redemptions may only be observed for the actually purchased brand, the price net of coupon is an endogenous variable. The effect of the endogeneity bias is to exaggerate price sensitivity. Because we do not know what coupons the consumer could have used if he/she had purchased another brand, the relative price of the purchased brand tends to be biased downward, making consumers appear more price sensitive. Incorporating coupons in the analysis would require explicit modeling of the coupon-redemption decision, which is beyond the scope of this study.

There were 19 stores with scanners in Sioux Falls during the sample period. Four of these had no price file because they were too small and are therefore excluded from the analysis. I concluded that two additional stores, although they did have price files, were sufficiently small that the price data would be unreliable [consisting almost entirely of imputations as in the preceding steps (2) and (3) rather than actual observed prices]. These were also excluded, leaving 13 stores.

The full dataset contained 2,320 consumers. To reduce the substantial computational burden involved in the estimation, I randomly selected consumers into the estimation sample with 50% probability. This produced roughly a half sample of 1,150 consumers with 5,353 purchase occasions. Thus, the mean number of purchase occasions per consumer is 4.65.

Attributes of alternatives included in the $\mathbf{A}$ vector are dummies for whether the brand is Hunt's or Del Monte, a dummy for whether the alternative comes in a plastic container, size and size squared in ounces, price, and dummies for whether a line, display, or major ad for the brand was in effect at the time of the purchase occasion. Characteristics of consumers included in $\mathbf{X}$ were household size and household income. However, $\boldsymbol{\beta}$ and ϕ_1 are restricted so that $\mathbf{X}_{it}\boldsymbol{\beta}_j$ is only a function of household size, but the interaction terms $\mathbf{A}_{jt}\mathbf{X}_{it}\phi_1$ for the price component of $\mathbf{A}$ involve both household size and income. A vector of ones is not included in $\mathbf{X}$, so there are no alternative-specific constants in the model. Thus, the model could be used to predict sales of a new size or package type for an existing brand (provided it can be positioned along the $\mathbf{L}$ and $\mathbf{P}$ vectors). The presence of brand intercepts precludes using the model to estimate sales of a completely new brand.

Finally, note that Heinz 32-ounce size is labeled alternative 7 and will be the base alternative in the MMP model. Thus, $\mathbf{L}_7 = \mathbf{P}_7 = 0$, and all positions of all other alternatives along the $\mathbf{L}$ and $\mathbf{P}$ vectors are relative to that of the Heinz 32-ounce size. Moreover, the attribute variables (i.e., price, size, plastic package, type of ad) are all measured relative to those for the Heinz 32-ounce size.

4. RESULTS

Tables 2–4 present results obtained by estimating models of successively greater complexity, to determine if the state-dependence parameters λ and α remain significant as more and more complex patterns of heterogeneity are included. The most straightforward models are presented in Table 2 because these contain heterogeneity structures similar to ones that have often been estimated in the literature. Tables 3 and 4 contain models with more complex heterogeneity distributions.

Model 1 in Table 2 is quite similar to that of Guadagni and Little (1983), except that the errors are iid normal instead of iid extreme value. The estimates of the ϕ_0 parameters (i.e., coefficients on observed attributes) all appear reasonable. Coefficients on the Hunt's and Del Monte dummies are negative and significant, indicating that, ceteris paribus, consumers prefer Heinz to these brands. The coefficient on the plastic dummy is positive and significant, indicating that consumers prefer plastic containers. The price coefficient is negative and significant as expected. The coefficients on the three types of ad dummies are all positive and significant. The estimates of the quadratic in package size indicate that consumers prefer smaller sizes, ceteris paribus. Because these patterns of coefficients on the attributes are consistent across all estimated model specifications, I will not discuss them further. (Estimates of the quadratic in package size do fluctuate substantially across specifications, but in all cases they indicate that consumers prefer smaller packages.)

The estimates of α and λ are .813 and 1.985, with standard errors of .015 and .106, respectively. These estimates imply that the lagged purchase has a strong effect on current decisions. Note that $\mathrm{GL}_{ijt} = \alpha\mathrm{GL}_{ij,t-1} + (1-\alpha)d_{ij,t-1}$. Thus, the increment in i's evaluation of the utility of purchasing j at t is $\lambda(1-\alpha)$ if i bought j at $t-1$ as opposed to a case in which i did not buy j at $t-1$. If we compare two identical consumers who face the same marketing mix and have identical histories except that consumer A chose alternative 1 last period but consumer B did not, the current-period utility evaluation of brand 1 will be roughly .371 greater for consumer A than consumer B. Given a price coefficient estimate of -1.453, this corresponds roughly to a 27-cent price cut for alternative 1 in terms of its impact on the probability that consumer A will buy alternative 1 in the current period.

Model 2 in Table 2 corresponds closely to a familiar Dirichlet multinomial-type model, except that Γ_{ij} is assumed to be normal rather than gamma. The parameter κ, which captures the degree of population heterogeneity in preferences for the unique type I attributes of brands, is estimated to be .915 with a standard error of .058. This implies a high degree of heterogeneity. In fact, the fraction of total error variance due to heterogeneous preferences for the type I unique factor is 47.8%.

Model 3 in Table 2 contains both κ heterogeneity and state dependence. The estimated strengths of both of these sources of persistence fall when they are included in the model simultaneously, relative to the cases in which each was included separately. The estimated fraction of total error variance due to the unique type I factor falls to 30.7%. If we again compare two consumers who face the same marketing mix and have identical histories except that consumer A chose alternative 1 last period while consumer B did not, the current-period utility evaluation of brand 1 will now be only .148 greater for consumer A than for con-

Table 2. The Interaction of Heterogeneity and State Dependence in Some Simple Models With Independent Time-Varying Error Components

Parameter name	Model 1: GL form of state dependence but no heterogeneity	Model 2: Kappa heterogeneity only and no state dependence (similar to Dirichlet multinomial)	Model 3: Kappa-type heterogeneity and GL form of state dependence	Model 4: Kappa heterogeneity, a common factor, and GL form of state dependence
Kappa		.915 (.058)	.444 (.052)	.221 (.040)
Lambda	1.985 (.106)		.889 (.154)	1.096 (.160)
Alpha	.813 (.015)		.833 (.029)	.836 (.026)
L1				−.361 (.084)
L2				−.531 (.102)
L3				.484 (.086)
L4				1.148 (.143)
L5				.897 (.115)
L6				.170 (.072)
Hunt's	−.428 (.023)	−.559 (.030)	−.526 (.028)	−.584 (.030)
Del Monte	−.644 (.030)	−.821 (.038)	−.763 (.036)	−.870 (.050)
Plastic	.330 (.023)	.347 (.030)	.373 (.028)	.391 (.029)
Ounces	−.024 (.005)	−.016 (.005)	−.023 (.005)	.001 (.008)
Oz.**2/100	.007 (.005)	−.009 (.006)	.000 (.006)	−.035 (.010)
Price	−1.453 (.053)	−1.663 (.057)	−1.656 (.058)	−1.695 (.067)
Line ad	.485 (.105)	.397 (.094)	.425 (.101)	.439 (.106)
Display ad	.718 (.043)	.583 (.038)	.638 (.042)	.698 (.044)
Major ad	.803 (.049)	.681 (.042)	.759 (.047)	.788 (.052)

NOTE: Standard errors follow the parameter estimates in parentheses.

sumer B. Given a price coefficient estimate of -1.656, this corresponds to only a 9-cent price cut for alternative 1 in terms of its impact on the probability consumer A will buy alternative 1 in the current period.

Model 4 in Table 2 has a heterogeneity distribution like that of Elrod and Keane (1995). It incorporates **L** heterogeneity by including a one-column **L** vector. This leads to a substantial reduction in the estimated degree of κ heterogeneity. Because the error variance for alternative j now depends on $\mathbf{L}_j$, the fraction of error variance due to the unique type I factor is alternative specific, but it never exceeds 18% and is as little as 11.8% for alternative 4 (the 64-oz. Heinz).

When **L** heterogeneity is included, the estimate of λ actually increases, but α is hardly affected. It may seem strange that the estimated strength of state dependence increases when an additional form of heterogeneity is included in the model. Intuitively, what is happening is that certain segments of consumers are observed to switch frequently among a small group of alternatives. A model with only κ heterogeneity and state dependence cannot explain such switching behavior. Therefore, the model downplays the strength of state dependence to capture the gross transition rate in the data (i.e., the probability of purchase of any brand other than j at t given that j was purchased at $t-1$). Once **L** heterogeneity is included, it is possible to have both a low gross transition rate and high transition rates among certain subsets of alternatives. This enables the model to reconcile frequent switching among subsets of alternatives with stronger state dependence.

The estimates of the elements of **L** imply that the alternatives 4 and 5 (the 64-oz. and 14-oz. Heinz) are perceived as most different from alternatives 1 and 2 (Hunt's and Del Monte). Thus, the unobserved type I common factor appears to be "Heinzness," with extremely large and small sizes of Heinz perceived as most "Heinz-like." Two unappealing features of this model are that Heinzness is not readily interpretable as a common factor and that it implies frequent consumer switching between the 64-ounce and 14-ounce sizes of Heinz. As we shall see later in Table 5 (p. 321), when a more flexible pattern of heterogeneity is allowed, the estimated elements of **L** change in such a way that the type I common factor appears to be package size—a much more appealing interpretation.

Given the results in Table 2, it would appear that both heterogeneity and state dependence explain a significant part of the persistence in brand choices. But are the GL parameters only significant because the functional form that I have assumed for heterogeneity is too simplistic? In Tables 3–4 I seek to answer this question by allowing for progressively more complex patterns of heterogeneity. In Table 3, results for models with no state dependence and only heterogeneity are reported. In Table 4 the results for the same models are reported, except that in each case the GL form of state dependence is included. Thus, we can determine if the GL parameters remain significant in models with each successively more complex pattern of heterogeneity.

Model 5 in Table 3 is an MMP model that allows the ε_{ijt} to have a one-factor covariance structure. In other words, a single type II common attribute is included by estimating a one-column **P** vector. This generalizes Model 2, which was an iid probit model. The estimated elements of the **P** vector are generally large and highly significant, indicating that independence of the time-varying preference shocks across brands is strongly rejected. This is not surprising, as we would expect time-varying preference shocks to be more

Table 3. Models With Heterogeneity/Serial Correlation and No State Dependence

Parameter name	Model 5: Let the ε_{ijt} be correlated across brands (P)	Model 6: Add correlation in random effects across brands (L)	Model 7: Let the ε_{ijt} be correlated over time, with AR(1) parameter rho	Model 8: Add observed heterogeneity in intercepts	Model 9: Add observed heterogeneity in price response	Model 10: Add unobserved heterogeneity in price response
Kappa	1.878 (.159)	1.211 (.102)	.971 (.104)	.950 (.101)	.949 (.098)	.913 (.100)
L1		−.504 (.127)	−.529 (.133)	−.535 (.163)	−.422 (.163)	−.513 (.265)
L2		−.990 (.171)	−.721 (.169)	−.893 (.164)	−.751 (.162)	−.684 (.169)
L3		.743 (.140)	1.218 (.182)	1.401 (.177)	1.391 (.195)	1.534 (.207)
L4		1.981 (.265)	2.187 (.298)	2.571 (.313)	2.736 (.346)	2.770 (.375)
L5		1.746 (.196)	1.110 (.172)	1.001 (.183)	.739 (.176)	.074 (.219)
L6		.437 (.147)	.678 (.157)	.744 (.141)	.801 (.169)	.849 (.142)
P1	.536 (.162)	.270 (.196)	.239 (.259)	.193 (.454)	.126 (.454)	.159 (.498)
P2	.705 (.211)	.425 (.221)	1.047 (.192)	1.112 (.199)	1.198 (.127)	1.291 (.245)
P3	.508 (.236)	−.086 (.160)	1.509 (.251)	1.686 (.190)	1.736 (.262)	1.964 (.225)
P4	1.727 (.273)	1.376 (.281)	1.736 (.387)	2.117 (.353)	2.323 (.345)	2.426 (.420)
P5	.484 (.282)	.446 (.417)	.130 (.377)	−.196 (.351)	−.380 (.169)	−.409 (.438)
P6	2.191 (.188)	1.834 (.173)	2.107 (.241)	2.222 (.253)	2.272 (.166)	2.278 (.219)
Rho			.257 (.027)	.257 (.029)	.257 (.027)	.262 (.030)
Hunt's	−.452 (.040)	−.561 (.045)	−.519 (.068)	−.526 (.122)	−.535 (.122)	−.515 (.139)
Del Monte	−.741 (.048)	−.999 (.068)	−.881 (.065)	−.902 (.134)	−.922 (.120)	−.885 (.130)
Plastic	.283 (.039)	.303 (.040)	.273 (.044)	.339 (.103)	.372 (.107)	.384 (.115)
Ounces	−.016 (.007)	.017 (.011)	.009 (.016)	−.046 (.019)	−.046 (.023)	−.045 (.024)
Oz.**2/100	−.008 (.009)	−.056 (.015)	−.048 (.020)	−.010 (.027)	−.014 (.031)	−.015 (.032)
Price	−1.613 (.071)	−1.865 (.084)	−1.765 (.090)	−1.782 (.086)	−1.744 (.190)	−1.743 (.212)
Price*inc					.010 (.003)	.010 (.004)
Price*HH size					−.150 (.053)	−.168 (.048)
Var(**P**)						1.001 (.183)
Line ad	.492 (.101)	.455 (.110)	.499 (.113)	.497 (.116)	.434 (.113)	.519 (.123)
Display ad	.613 (.041)	.650 (.045)	.722 (.048)	.744 (.049)	.745 (.048)	.723 (.049)
Major ad	.765 (.048)	.814 (.053)	.859 (.057)	.885 (.056)	.875 (.059)	.903 (.056)
HH size-1				−.019 (.025)	−.029 (.025)	−.032 (.026)
HH size-2				−.024 (.033)	−.026 (.033)	−.028 (.033)
HH size-3				.022 (.029)	.035 (.032)	.029 (.036)
HH size-4				.079 (.055)	.075 (.057)	.089 (.058)
HH size-5				−.263 (.037)	−.268 (.050)	−.313 (.054)
HH size-6				−.070 (.028)	−.070 (.034)	−.056 (.034)

NOTE: Standard errors follow the parameter estimates in parentheses.

similar for brands that are closer together in the attribute space.

In Model 6, **L** heterogeneity is included by estimating a one-column **L** vector. In Model 7, the permanent–transitory distinction is relaxed by allowing the time-varying part of the errors to follow an AR(1) process. Here, the AR(1) parameter ρ is estimated to be .257 with a standard error of .027. Thus, equicorrelation is clearly rejected.

Comparison of Models 5, 6, and 7 show how the importance of κ heterogeneity is seriously overestimated by models that ignore other types of heterogeneity. The estimate of κ falls from 1.878 to 1.211 when L heterogeneity is included, and then to .971 when persistence in the time-varying error component is included. This suggests that models like the Dirichlet multinomial may seriously overestimate the degree to which consumers are intractably loyal to their favorite brands.

In Models 8–10, the effects of including observed heterogeneity in intercepts and both observed and unobserved heterogeneity in slopes are considered. In Model 8, the $\mathbf{X}_{it}\boldsymbol{\beta}_j$ term in Equation (1) is brought in by letting preferences among alternatives depend on household size. The household-size coefficient for alternative 5, the 14-ounce Heinz, is $-.263$ with a standard error of .037. This indicates, as we would expect, that larger households typically dislike small package sizes. Model 9 brings in the $\mathbf{A}_{jt}\mathbf{X}_{it}\phi_1$ term by letting the price coefficient differ by household income and household size. The coefficients on the price with income and household size interactions are .010 and $-.150$, with standard errors of .003 and .053, respectively. Because income is measured in thousands of dollars, these estimates imply that a $15,000 increase in household income increases the price coefficient by .15, which is roughly the same as the effect of having one fewer household member. Model 10 brings in the $A_{jt}\nu_i$ term by letting the price coefficient be normally distributed in the population. The estimated variance σ_ν^2 is quite large at 1.001 with a standard error of .183.

The introduction of the additional forms of heterogeneity in Models 8–10 has little effect on the estimated importance of κ heterogeneity. But there is an interesting effect on the estimated **L** vector. As additional forms of heterogeneity are included, alternative 5, the 14-ounce Heinz, moves from a large positive loading to a position near 0. Because the two largest sizes, alternatives 4 and 3, now have the two largest loadings, the underlying type I common factor begins to look like package size, but the map continues to place alternatives 5 and 6 somewhat out of proper

Table 4. Models With Heterogeneity/Serial Correlation and State Dependence

Parameter name	Model 11: Let the ε_{ijt} be correlated across brands (P)	Model 12: Add correlation in random effects across brands (L)	Model 13: Let the ε_{ijt} be correlated over time, with AR(1) parameter rho	Model 14: Add observed heterogeneity in intercepts	Model 15: Add observed heterogeneity in price response	Model 16: Add unobserved heterogeneity in price response
Kappa	.646 (.106)	.529 (.089)	.431 (.081)	.475 (.091)	.501 (.091)	.549 (.092)
Lambda	1.456 (.193)	1.120 (.214)	1.355 (.336)	1.433 (.426)	1.391 (.356)	1.346 (.379)
Alpha	.850 (.021)	.868 (.030)	.897 (.034)	.911 (.033)	.907 (.033)	.909 (.037)
L1		−.342 (.110)	−.246 (.106)	−.255 (.106)	−.248 (.113)	.090 (.103)
L2		−.854 (.147)	−.406 (.129)	−.139 (.144)	−.028 (.155)	.303 (.151)
L3		.707 (.123)	1.180 (.180)	1.490 (.190)	1.664 (.211)	1.857 (.237)
L4		1.876 (.317)	2.007 (.323)	2.550 (.378)	2.599 (.368)	2.454 (.320)
L5		1.385 (.213)	1.129 (.167)	.669 (.167)	.699 (.178)	−1.959 (.328)
L6		.434 (.134)	.637 (.141)	.826 (.144)	.976 (.158)	1.285 (.192)
P1	1.419 (.224)	.544 (.179)	.361 (.196)	.317 (.233)	.434 (.210)	.276 (.168)
P2	2.041 (.314)	.540 (.228)	.976 (.229)	1.621 (.265)	1.755 (.259)	1.280 (.198)
P3	−1.224 (.128)	−.276 (.154)	1.871 (.283)	2.211 (.270)	2.456 (.292)	2.796 (.307)
P4	1.663 (.302)	2.047 (.371)	2.092 (.444)	2.670 (.407)	2.776 (.407)	2.543 (.385)
P5	1.124 (.252)	1.183 (.340)	−.241 (.311)	−.731 (.380)	−.653 (.387)	−.902 (.361)
P6	3.345 (.323)	1.841 (.203)	1.946 (.236)	2.118 (.344)	2.341 (.253)	2.804 (.311)
Rho			.139 (.032)	.171 (.031)	.167 (.031)	.134 (.033)
Hunt's	−.202 (.045)	−.492 (.049)	−.515 (.060)	−.584 (.105)	−.470 (.110)	−.465 (.099)
Del Monte	−.581 (.069)	−.916 (.076)	−.778 (.069)	−.848 (.134)	−.715 (.141)	−.692 (.121)
Plastic	.147 (.042)	.322 (.042)	.454 (.048)	.618 (.104)	.569 (.119)	.572 (.139)
Ounces	−.102 (.009)	.005 (.018)	.004 (.023)	−.058 (.026)	−.061 (.028)	−.042 (.032)
Oz.**2/100	.015 (.011)	−.050 (.026)	−.046 (.029)	−.009 (.036)	−.001 (.037)	−.015 (.038)
Price	−1.943 (.118)	−1.895 (.097)	−2.052 (.105)	−2.056 (.111)	−1.898 (.241)	−2.355 (.261)
Price*inc					.011 (.003)	.023 (.005)
Price*HH size					−.143 (.062)	−.206 (.072)
Var(**P**)						1.818 (.314)
Line ad	.574 (.114)	.643 (.125)	.632 (.130)	.626 (.135)	.633 (.136)	.582 (.140)
Display ad	.650 (.052)	.720 (.052)	.761 (.055)	.778 (.059)	.777 (.060)	.800 (.055)
Major ad	.834 (.063)	.888 (.060)	.975 (.066)	1.011 (.067)	1.031 (.069)	1.042 (.068)
HH size-1				−.011 (.023)	−.023 (.023)	−.041 (.025)
HH size-2				.018 (.032)	−.000 (.032)	−.009 (.032)
HH size-3				.008 (.034)	.029 (.039)	−.012 (.044)
HH size-4				.103 (.059)	.130 (.060)	.143 (.058)
HH size-5				−.348 (.056)	−.330 (.076)	−.607 (.077)
HH size-6				−.089 (.029)	−.061 (.035)	−.073 (.040)

NOTE: Standard errors follow the parameter estimates in parentheses.

position. Interestingly, a high correlation of alternative positions along the **L** and **P** vectors is also apparent. This is consistent with a story in which both the underlying type I and type II common factors are package size. Consumers have a baseline preference for a certain size (based perhaps on how much their family members like ketchup), but they also go through phases in which they may prefer a larger size or smaller size (perhaps because they go through phases in which they have a lot of cookouts or eat a lot of hamburgers, followed by phases in which they tire of these activities).

Table 4 presents results analogous to those in Table 3, except that the GL form of state dependence is included in each model. Comparing Model 11 in Table 4 with Model 5 in Table 3, we see that introduction of state dependence leads to a dramatic reduction in the estimated degree of κ heterogeneity. The estimate of κ falls from 1.878 to .646. This comparison even understates the decline, because the length of the **P** vector also increases substantially when state dependence is included. Thus, the fraction of total error variance due to the type I unique factor is in a range of 35.6% to 62.7% in Model 5 but in a range of only 8.9% to 23.3% in Model 11.

It is also interesting to compare Model 11 with Model 3 in Table 2. This isolates the effect of including the **P** vector in a model with both κ heterogeneity and state dependence. Note that κ increases from .444 to .646 while λ increases from .889 to 1.456. The estimated value of $\mathbf{P}_6$ in Model 11 is quite large (3.345) implying frequent switching in and out of alternative 6 (Heinz 28-oz. plastic) over time. Furthermore, the estimated value of $\mathbf{P}_3$ is a large negative (−1.224), implying frequent switching between alternatives 6 and 3 (Heinz 40-oz. plastic). Once the model is free to capture this pattern of switching with a large positive $\mathbf{P}_6$ value and a large negative $\mathbf{P}_3$ value, it can reconcile the observed data with stronger κ heterogeneity and state dependence.

To get some idea of the estimated strength of state dependence in Model 11, note that the estimates of α and λ are .850 and 1.456, respectively. If we compare two consumers who face the same marketing mix and have identical histories, except that consumer A chose alternative 1 last period while consumer B did not, the current-period utility evaluation of brand 1 will be roughly .218 greater for A than for B. Given a price coefficient of −1.943, this corresponds to roughly an 11-cent price cut for alternative 1 in terms of its impact on the probability consumer A will buy alternative

1 in the current period. The comparable figure was 9 cents in Model 3 in Table 2.

In Model 12 in Table 4, a single column of **L** is included. As compared to Model 11, this causes the estimated value of λ to fall from 1.456 to 1.120 and the estimated value of κ to fall from .646 to .529. The drop in λ appears to contradict the finding in Table 2 that λ increased when **L** was included. Apparently, once certain types of switching behavior are accounted for by inclusion of the **P** vector, the additional inclusion of **L** heterogeneity does indeed lead to a decline in the estimated strength of state dependence. Moreover, comparison of Model 12 with Model 3 indicates that inclusion of **L** and **P** together does lead to an increase in the estimated strength of state dependence.

In Model 13 the equicorrelation assumption is relaxed by including the ρ parameter. The estimate of ρ is .139 with a standard error of .032. Comparison with Model 7 indicates that inclusion of state dependence reduces the estimate of ρ almost in half. Inclusion of ρ has little effect on the estimates of the GL parameters. Comparing Models 12 and 13, we see that the estimate of λ increases from 1.120 to 1.355 while the estimate of α increases from .868 to .897. Thus, $\lambda(1-\alpha)$ drops from .148 to .140. This means that last period's purchase has only slightly less impact on current-period utility evaluations. Intuitively, when ρ is included the model ascribes some of the short-run persistence in choice behavior to short-run persistence in preference shocks. This leads it to ascribe slightly less of the observed persistence to state dependence, but the effect is minor.

In Models 14–16 we see the effects of including observed heterogeneity in intercepts and both observed and unobserved heterogeneity in slopes. In Model 14, preferences among alternatives are allowed to depend on household size. In Model 15 the price coefficient is allowed to differ by household income and household size, and in Model 16 the price coefficient is allowed to be normally distributed in the population. None of these changes lead to substantial diminution in the estimated strength of state dependence.

The bottom-line results are the estimates of Model 16. This "full model" contains the GL parameters along with all seven types of heterogeneity. Despite all these controls for heterogeneity, the parameters capturing state dependence remain highly significant. The estimate of λ is 1.346 with a standard error of .379, while the estimate of α is .909 with a standard error of .037. Based on the point estimates for Model 16, if we compare two consumers who face the same marketing mix and have identical histories except that consumer A chose alternative 1 last period while consumer B did not, the current-period utility evaluation of brand 1 will be roughly .122 greater for consumer A than for consumer B. Because a typical family has a price coefficient of about -2.4, these will have effects similar to roughly a 5-cent price cut at time t. The effects on utility evaluations at $t+1$ through $t+10$ will be .111, .101, .092, .084, .076, .069, .063, .057, .052, and .047, respectively. The point estimates from Model 1—the GL-type model—imply a corresponding current-period effect of .371 (comparable to a 27-cent price cut) and $t+1$ through $t+10$ effects of .302, .245, .199, .162, .132, .107, .087, .071, .058, and .047. Thus, as expected, the model with all seven types of heterogeneity implies a much smaller short-run effect of past purchases on current utility evaluations. Both models imply rather small effects by period $t+10$.

Other parameter estimates in Model 16 are also of interest. The estimate of κ is .549 with a standard error of .092. This, combined with the estimates of **L**, **P**, ρ, and σ_ν^2, implies that the fraction of total error variance due to κ heterogeneity is in the range of 3.9% (alternative 4) to 22.1% (alternative 1). In Model 10, which has all seven types of heterogeneity but no state dependence, the estimate of κ is .913 and the fraction of total error variance due to κ heterogeneity is in the range of 9.0% (alternative 4) to 38.3% (alternative 5). Thus, allowing for state dependence substantially reduces the estimated degree of κ heterogeneity.

The estimates of the elements of the **L** vector indicate that alternatives 4 and 5 are at opposite ends of the one-dimensional map, with 3 positioned close to 4, alternatives 1, 2, and 7 near the middle, and alternative 6 between 2 and 3. Alternatives 4 and 5 are the 64-ounce and 14-ounce sizes, alternative 3 is the 40-ounce size, alternatives 1, 2, and 7 are 32-ounce sizes, and alternative 6 is a 28-ounce size. Thus, in terms of position on the **L** vector, all alternatives except 6 line up perfectly by size, and the unobserved type I common factor can be plausibly interpreted as package size. Note that I am referring to population heterogeneity in preferences for package size that are not captured by the observed differences in household size. In Model 10, alternative 5 (Heinz 14-oz.) was placed at the center of the map. It is interesting that the estimated market map is very sensitive to inclusion of state dependence and that inclusion of state dependence is necessary before the market map takes on an easily interpretable form [see Erdem (1996) for a thorough discussion of this issue].

In Model 16 the price coefficient is -2.355 with a standard error of .314, but the coefficients on the price with income and household size interactions are .023 and $-.206$, with standard errors of .005 and .072, respectively. These estimates imply, for example, that a typical household with an income of $30,000 and three household members will have a price coefficient of -2.283. The estimated population variance in the price coefficient is substantial at 1.818, implying a standard deviation of 1.348. Comparison with Model 10 reveals that the inclusion of state dependence leads to increases in the estimated degree of both observed and unobserved heterogeneity in the price coefficient. This highlights again the point that proper segmentation of a market by consumer position in the heterogeneity distribution will in general require control for the presence of state dependence.

5. ROBUSTNESS CHECKS

The full model estimated in Section 4 allows for true state dependence of the GL form, along with seven types of heterogeneity that can also lead to serial persistence in choices. My main conclusion is that, even after allowing for all of these nonstate-dependence-based sources of serial persistence in the model, the GL form of state dependence remains significant. Thus, it appears likely that true state dependence is present in the Nielsen ketchup choice data.

Table 5. Robustness Checks on General Model

Parameter name	Model 16: Full model	Model 17: Let AR(1) parameters differ by brand	Model 18: Include 2nd column of L matrix	Model 19: Subsample with at least 6 periods of data
Kappa	.549 (.092)	.544 (.091)	.577 (.101)	.580 (.097)
Lambda	1.346 (.379)	1.341 (.389)	1.271 (.388)	1.348 (.411)
Alpha	.909 (.037)	.908 (.038)	.908 (.036)	.897 (.041)
L11	.090 (.103)	.092 (.106)	.081 (.099)	.031 (.164)
L12	.303 (.151)	.298 (.151)	.269 (.671)	.347 (.237)
L13	1.857 (.237)	1.834 (.242)	1.766 (.785)	1.887 (.227)
L14	2.454 (.320)	2.453 (.335)	2.439 (.829)	2.625 (.377)
L15	−1.959 (.328)	−1.944 (.301)	−1.882 (.515)	−2.245 (.407)
L16	1.285 (.192)	1.289 (.200)	1.288 (.736)	1.307 (.257)
L22			−.117 (.986)	
L23			.358 (5.577)	
L24			.293 (7.914)	
L25			.092 (5.257)	
L26			.022 (4.043)	
P1	.276 (.168)	.279 (.170)	.263 (.121)	.284 (.289)
P2	1.280 (.198)	1.277 (.200)	1.266 (.225)	1.279 (.242)
P3	2.796 (.307)	2.789 (.312)	2.740 (.330)	2.882 (.347)
P4	2.543 (.385)	2.540 (.403)	2.542 (.462)	2.584 (.426)
P5	−.902 (.361)	−.886 (.371)	−.804 (.416)	−.793 (.614)
P6	2.804 (.311)	2.801 (.320)	2.772 (.274)	2.930 (.515)
Rho or Rho-1	.134 (.033)	.179 (.071)	.135 (.032)	.142 (.031)
Rho-2		.100 (.060)		
Rho-3		.162 (.041)		
Rho-4		.153 (.053)		
Rho-5		.191 (.155)		
Rho-6		.108 (.034)		
Hunt's	−.465 (.099)	−.464 (.093)	−.464 (.107)	−.462 (.144)
Del Monte	−.692 (.121)	−.690 (.120)	−.697 (.133)	−.687 (.200)
Plastic	.572 (.139)	.569 (.137)	.557 (.140)	.588 (.240)
Ounces	−.042 (.032)	−.042 (.033)	−.042 (.032)	−.039 (.049)
Oz.**2/100	−.015 (.038)	−.015 (.039)	−.015 (.038)	−.013 (.053)
Price	−2.355 (.261)	−2.337 (.247)	−2.272 (.274)	−2.446 (.330)
Price*inc	.023 (.005)	.023 (.005)	.022 (.005)	.020 (.007)
Price*HH size	−.206 (.072)	−.206 (.071)	−.209 (.071)	−.224 (.100)
Var(**P**)	1.818 (.314)	1.814 (.312)	1.804 (.329)	1.998 (.384)
Line ad	.582 (.140)	.577 (.140)	.546 (.140)	.494 (.186)
Display ad	.800 (.055)	.796 (.054)	.790 (.060)	.832 (.059)
Major ad	1.042 (.068)	1.040 (.067)	1.045 (.072)	.986 (.078)
HH size-1	−.041 (.025)	−.041 (.025)	−.041 (.024)	−.045 (.042)
HH size-2	−.009 (.032)	−.008 (.032)	−.009 (.032)	−.010 (.051)
HH size-3	−.012 (.044)	−.011 (.043)	−.007 (.046)	−.013 (.061)
HH size-4	.143 (.058)	.143 (.057)	.145 (.060)	.125 (.077)
HH size-5	−.607 (.077)	−.603 (.068)	−.580 (.093)	−.674 (.066)
HH size-6	−.073 (.040)	−.073 (.039)	−.073 (.041)	−.078 (.056)

NOTE: Standard errors follow the parameter estimates in parentheses.

Although the heterogeneity structure that I have estimated is quite flexible, one might still argue that some of the restrictions I imposed in estimation are generating the significance of the state-dependence parameters. This section investigates the robustness of my results to certain changes in specification of the full model—Model 16. First, to conserve on parameters, in Model 16, I restricted the AR(1) parameters for all alternatives to be identical. In Model 17 in Table 5, I relax this restriction and allow a separate AR(1) parameter for each alternative. Recall that in Model 16 the estimate of the AR(1) parameter was .134 with a standard error of .033. In Model 17 the estimated AR(1) parameters range from .100 to .191. In no case is an estimate even one standard deviation away from .134. Furthermore, comparison of Models 16 and 17 indicates that the other model parameters scarcely change. In particular, the coefficient multiplying the state-dependence variable only changes from 1.346 to 1.341.

Another key way in which I conserved on parameters in Model 16 was by imposing a factor structure on the error covariance matrix. If utility for each alternative was allowed to depend on a random effect, with the vector of random effects having an unrestricted covariance matrix, there would be a 6×6 covariance matrix with 21 unique elements to be estimated. By imposing the $\kappa + 1$ column of L structure (i.e., each alternative has a unique factor and one common factor), I reduce the number of unique elements in the covariance matrix of the random effects to 7. Are results sensitive to this restriction? Model 18 in Table 5 addresses this question by including a second column in the **L** matrix.

When the second column of **L** is included, the model parameters other than the common factor loadings (elements of **L**) are little affected. In particular, the coefficient multiplying the state-dependence variable only changes from 1.346 to 1.271. Note also that the standard errors for the elements of the first column of $\mathbf{L}$ increase substantially. The standard errors for the elements of the second column of $\mathbf{L}$ are enormous, and none of the elements of the second column approach significance. All of these results are consistent with earlier Monte Carlo work by Keane (1992) and Geweke et al. (1994), which indicates that it is very difficult to uncover unrestricted covariance structures for high-dimensional choice models. Basically, discrete choice data does not contain very much information about error covariances. As a consequence, it would be very unusual to reject the restriction that the covariance matrix for a high-dimensional choice model lies in a rather low-dimensional factor space.

A final important question is whether my results are sensitive to my ad hoc specification of the initial conditions. As I discussed in Section 4, with an ad hoc specification of the initial conditions, consistency will only be achieved as $T \to \infty$ because as $T \to \infty$ the specification of the initial conditions becomes less important. This suggests that a way to test the sensitivity of the results to specification of the initial conditions is to reestimate the model on a subsample of individuals who have relatively long strings of observed purchases. Recall that the full sample has 1,150 people with 5,353 purchase occasions. Model 19 in Table 5 is the same as Model 16 except that it is estimated on a subsample of 329 people who have at least six observed purchases. This subsample has a total of 3,266 purchase occasions—an average of 9.93 per person, as compared to only 4.65 in the full sample. Comparison of Models 19 and 16 indicates that parameter estimates are very similar in the subsample. In particular, the coefficient multiplying the state-dependence variable only changes from 1.346 to 1.348. This suggests that results are not too sensitive to the specification of initial conditions. I think the reason this is true is that most of the data points in the full sample are for people with relatively large numbers of purchase occasions (note that the subsample used to estimate Model 19 has only 28.6% of the people but 61.0% of the data points). Hence, most of the information for identification of the serial correlation and state-dependence parameters is coming from people with rather long purchase strings. Overall, I conclude that the finding that true state dependence is present in the data is remarkably robust to alternative specifications of the model.

6. MODEL FIT EVALUATION

Although the results in Sections 4 and 5 indicate that parameters capturing true state dependence remain significant in the model even after many other nonstate-dependence-based sources of serial persistence are controlled for, an important question is: "How much better does the full model that includes state dependence fit than the other models?" Because the models are estimated using MSM, there is no omnibus fit statistic available, like the likelihood function, that can be used to compare fit across models. Instead, I look at how four selected models fit various transition rates in the data.

I consider one-, two-, and three-period transition matrices. The one-period transition matrix is the 7×7 matrix that contains the values $\sum_{i=1,N} \sum_{t=1,T} d_{ijt} | d_{ik,t-1} = 1$ for $j = 1, 7$, $k = 1, 7$. A two-period transition matrix is a 7×7 matrix that contains $\sum_{i=1,N} \sum_{t=1,T} d_{ijt} | d_{ik,t-1} = 1$, $d_{ik,t-2} = q$ for $j = 1, 7$, $k = 1, 7$. There are two such matrices, depending on whether $q = 1$ or 0. A three-period transition matrix is a 7×7 matrix that contains $\sum_{i=1,N} \sum_{t=1,T} d_{ijt} | d_{ik,t-1} = 1$, $d_{ik,t-2} = q$, $d_{ik,t-3} = r$ for $j = 1, 7$, $k = 1, 7$. There are four such matrices, depending on whether q and r are set equal to 0 or 1.

For the Sioux Falls data, the one-period transition matrix is given in the upper left cell of Table 6. The rightmost column of the table gives the row sums (number of occasions that each alternative was bought at $t-1$), and the bottom row gives the column sums (number of occasions that each alternative was bought at t). The two-period transition matrix that conditions on $d_{ik,t-1} = 1$, $d_{ik,t-2} = 1$ is in the middle column, and the three-period transition matrix that conditions on $d_{ik,t-1} = 1$, $d_{ik,t-2} = 1$, $d_{ik,t-3} = 1$ is in the rightmost column.

Note that the ketchup data is not characterized by a particularly high degree of persistence in choices at the aggregate level. For example, of the 1,495 cases of people buying alternative 6 (the Heinz 28-oz. plastic) at time $t-1$, in only 476 instances does the person buy it again on purchase occasion t—a 31.8% repeat-purchase probability. A person who buys alternative 3 on occasion $t-1$ (the Heinz 40-oz. plastic) is actually more likely to buy alternative 6 on the next occasion than to stay with 3. (Note that this switching pattern was picked up by the estimated models—see Sec. 4.)

Persistence appears to be greater when one conditions on a person having bought the same alternative on two previous purchase occasions. For example, of the 340 cases of people buying alternative 6 (the Heinz 28-oz. plastic) at both time $t-1$ and $t-2$, the person buys it again 184 times on purchase occasion t—a 54.1% repeat-purchase probability (versus 31.8% when conditioning only on $t-1$). And for the 139 cases of people buying alternative 6 at $t-1$, $t-2$, and $t-3$, the person buys it again 81 times on purchase occasion t—a 60.4% repeat-purchase probability.

To compare model fit, predicted one-, two-, and three-period transition matrices were constructed for four key models—Model 1, the GL-type model; Model 2, the Dirichlet multinomial-like model with κ heterogeneity only; Model 10, the model with the very general heterogeneity structure but no true state dependence; and Model 16, the full model that includes a very general heterogeneity structure plus true state dependence. Transition matrices obtained from simulation of the four models are reported in Table 6. Visual inspection of the one-period transition matrices reveals that Model 1 (GL) overstates persistence for the two largest market-share alternatives (6 and 7) while understating it for the smaller alternatives. Model 2 (κ heterogeneity) understates persistence for all alternatives except 3 and 6. Models 10 and 16 both look better, and it is difficult to choose between them solely from visual impression.

Turning to the two-period transition matrices, we see that Model 1 even more strikingly exaggerates persistence. For example, it predicts 570 events in which consumers buy alternative 6 at both $t-1$ and $t-2$, followed by 414 events in which they buy 6 again at time t—a repurchase rate of

Table 6. Predicted Versus Actual Transition Rates

Models	One-period transition frequencies: $Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1)$								Two-period transition frequencies: $Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1, d_{ij,t-2} = 1)$								Three-period transition frequencies: $Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1, d_{ij,t-2} = 1)$ $d_{ij,t-3} = 1)$							
Actual	263	46	32	4	4	155	143	647	116	8	6	0	1	39	31	201	61	4	2	0	0	14	9	90
Sioux Falls	69	68	6	6	5	49	84	287	14	27	1	0	0	7	12	61	1	17	0	0	0	3	3	24
data	42	11	85	26	9	107	75	355	8	0	23	9	2	21	8	71	3	0	6	3	0	6	2	20
	9	5	23	60	3	19	25	144	0	1	5	36	0	3	5	50	0	0	3	24	0	1	1	29
	10	1	13	2	50	30	27	133	1	0	1	0	24	6	3	35	0	0	0	0	12	4	0	16
	170	32	114	23	24	476	303	1,142	40	5	34	5	5	184	67	340	14	1	13	1	2	81	27	139
	171	84	74	26	20	363	757	1,495	40	23	23	11	5	116	417	635	15	5	6	5	1	61	270	363
	734	247	347	147	115	1199	1414	4,203	219	64	93	61	37	376	543	1,393	94	27	30	33	15	170	312	681
Model 1:	204	37	47	21	18	134	153	615	67	10	13	8	5	20	30	154	27	3	1	4	1	4	9	51
GL form	32	39	16	11	10	62	64	235	3	8	1	2	2	5	10	32	1	1	0	1	1	1	2	7
of state	43	8	56	9	5	88	79	289	3	2	18	1	1	7	12	43	1	0	9	0	0	2	3	17
dependence	16	10	7	17	3	30	31	114	1	2	2	3	0	1	2	11	0	0	1	0	0	0	1	2
but no	21	8	11	3	14	30	33	121	2	1	1	0	1	1	1	7	0	0	0	0	0	0	0	1
heterogeneity	139	52	62	27	23	729	270	1,303	39	12	14	10	4	414	76	570	16	6	9	6	1	271	20	330
	148	44	83	30	33	296	890	1,526	52	11	28	11	9	109	508	727	17	6	12	5	5	49	315	409
	603	200	284	121	107	1368	1520	4,203	168	46	78	35	22	557	638	1,544	64	17	34	16	8	328	350	817
Model 2:	224	39	69	18	16	135	148	650	71	13	24	4	5	21	33	172	31	2	9	1	2	5	4	55
Kappa-type	28	48	15	9	15	62	63	240	1	11	1	4	3	8	8	36	0	4	0	1	2	0	1	9
heterogeneity	50	13	111	10	11	89	84	369	9	2	44	2	2	14	14	86	3	1	25	1	1	5	3	38
only and no	14	11	11	30	6	29	22	123	0	0	2	10	0	3	4	20	0	0	2	2	0	1	2	9
state	20	9	17	7	33	48	45	180	1	0	2	2	8	4	6	24	1	0	0	1	1	2	2	7
dependence	153	52	89	35	46	598	293	1,267	44	13	24	11	14	271	84	463	17	10	11	4	4	144	20	211
	169	50	79	32	55	331	657	1,374	59	16	29	10	24	107	300	546	20	7	9	4	9	32	155	238
	659	224	393	143	184	1290	1310	4,203	187	57	126	44	58	428	448	1,347	74	25	57	14	20	189	188	567
Model 10:	276	54	39	8	13	139	173	701	91	19	6	3	3	37	46	204	41	4	1	1	1	10	8	67
Kappa-type	51	52	8	2	5	35	78	231	7	12	1	1	1	4	13	39	1	6	0	0	1	1	0	11
heterogeneity	31	6	110	16	14	84	82	343	6	2	34	6	7	13	14	83	2	0	11	3	3	5	2	27
and random	4	2	11	67	6	33	24	147	1	1	1	34	2	11	7	57	1	1	1	19	1	8	3	33
effects and	16	5	13	4	36	40	22	137	3	0	2	1	6	10	3	26	1	0	0	0	0	2	0	4
AR(1) error	149	36	86	35	28	547	304	1,186	44	9	19	11	11	226	90	412	16	3	9	3	3	96	41	170
but no state	194	68	85	29	37	314	730	1,458	69	31	40	11	14	122	319	608	25	16	14	8	5	51	135	255
dependence	722	223	354	162	140	1191	1411	4,203	221	76	104	68	45	423	492	1,429	87	31	38	35	13	172	190	567
Model 16:	245	51	43	13	12	142	151	657	87	12	12	4	3	37	32	187	36	6	4	1	1	7	6	61
Full model	59	59	9	6	6	43	83	266	11	15	1	1	1	4	12	45	3	7	0	0	1	0	2	14
	46	9	136	16	9	73	73	363	10	2	68	5	2	9	14	110	6	1	42	4	1	5	2	61
	9	4	16	37	2	29	30	127	1	1	7	12	1	4	4	30	0	0	2	5	0	2	2	12
	12	3	7	2	77	30	21	152	1	1	3	1	33	8	5	52	1	0	0	0	17	1	2	23
	136	52	87	36	27	564	294	1,197	37	13	24	14	6	247	84	426	16	6	8	2	1	117	41	191
	171	71	72	32	31	301	761	1,441	60	23	32	9	11	106	385	627	18	12	12	6	4	54	200	307
	679	251	372	144	164	1180	1412	4,203	208	67	148	47	57	415	536	1,477	81	33	70	19	23	187	255	669

NOTE: For the four estimated models, simulations were performed to generate unconditional choice sequences. The upper left 7*7 block of each matrix contains predicted transition frequencies, while the right-most column contains row sums and the bottom row contains column sums. Expected choice frequencies are rounded to the nearest integer value. Therefore row and column entries may not add up to the reported row and column sums.

72.6%, which is much too high. Model 2 continues to overstate persistence for 3 and 6 and understate it for other alternatives. Again models 10 and 16 look much better, and now a close visual inspection reveals some points of advantage for Model 16 over Model 10. In particular, model 10 substantially understates persistence for alternative 7.

The advantage of Model 16 over all other models becomes much more apparent when the three-period transition matrices are examined. Most obviously, while the three-period transition matrix for the actual Sioux Falls data has a grand sum over all rows and columns of 681, that for Model 1 is 817 and that for Models 2 and 10 is only 567. So Model 1 substantially overstates the frequency with which consumers purchase the same alternative on three successive occasions, but Models 2 and 10 substantially understate this frequency. Model 16 comes in just about right at 669.

The visual impression of superior performance of Model 16 in capturing transition rates is confirmed by examination of formal χ^2 test statistics for fit of the predicted to the actual cell counts in each transition matrix. These are reported in Table 7. Here I report results not just for the three transition matrices reported in Table 6 but also for the two-period transition matrix conditional on the $(d_{ik,t-1}, d_{ik,t-2}) = (1,0)$ pattern and the three-period transition matrices conditional on the $(d_{ik,t-1}, d_{ik,t-2}, d_{ik,t-3}) = (1,1,0)$, $(1,0,1)$ and $(1,0,0)$ patterns. Note that Model 16 clearly dominates Model 10 according to these formal χ^2 fit tests, even for the one-period transition rates in which its superiority was not visually obvious in Table 6. Model 16 is especially superior for the transition matrix conditional on the $(d_{ik,t-1}, d_{ik,t-2}, d_{ik,t-3})$ sequence, where a value of 247.7 is obtained versus 723.2 for Model 10. Apparently, the ad-

Table 7. Fit of Alternative Models to Transition Rates

Transition matrix	Model 1: GL form of state dependence but no heterogeneity	Model 2: Kappa-type heterogeneity only and no state dependence	Model 10: Kappa-type heterogeneity, random effects, and AR(1) error component but no state dependence	Model 16: Full model
	One-period transition rates			
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1)$	682.1	396.9	180.6	169.6
	Two-period transition rates			
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1,\ d_{ij,t-2} = 1)$	1,544.7	483.3	245.5	199.3
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1,\ d_{ij,t-2} = 0)$	312.3	217.3	233.2	145.1
	Three-period transition rates			
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1,\ d_{ij,t-2} = 1,\ d_{ij,t-3} = 1)$	4,885.8	639.2	723.2	247.7
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1,\ d_{ij,t-2} = 1,\ d_{ij,t-3} = 0)$	282.3	174.8	83.1	85.1
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1,\ d_{ij,t-2} = 0,\ d_{ij,t-3} = 1)$	214.3	159.4	104.5	98.6
$\Pr(d_{ijt} = 1 \mid d_{ij,t-1} = 1,\ d_{ij,t-2} = 0,\ d_{ij,t-3} = 0)$	260.7	152.5	142.0	100.3

NOTE: The numbers in the table are chi-squared statistics for fit of the models to the 7*7 transition matrices indicated in the leftmost column. The chi-squared statistics are defined as the sum over transition matrix elements of the squared difference between the actual and predicted transition count divided by the predicted transition count.

dition of the GL form of state dependence to the model greatly improves its ability to fit transition rates, even when many types of heterogeneity and serial correlation are allowed for.

7. POLICY SIMULATIONS

The various models estimated in Section 4 have very different implications with regard to the effect of changes in the marketing mix on dynamic consumer behavior. This section examines the impact of a one-period price promotion on long-term behavior to illustrate this point.

Table 8. Frequency Distribution of Purchases by Purchase Occasion

Purchase occasion	Alternative 1	2	3	4	5	6	7
1	151	80	91	29	72	331	396
2	133	71	93	15	42	264	288
3	116	53	52	31	23	213	212
4	89	27	59	16	23	158	178
5	70	13	34	18	5	144	142
6	69	14	19	11	7	85	124
7	57	13	22	11	5	65	85
8	40	8	12	9	2	55	72
9	28	14	10	6	0	41	60
10	22	8	10	8	3	41	39
11	16	8	10	5	0	21	44
12	18	6	8	4	1	19	35
13	9	3	4	1	1	20	24
14	8	3	3	1	1	13	25
15	5	1	3	2	1	9	18
16	4	2	0	0	0	12	17
17	5	0	1	2	0	11	11
18	3	0	3	1	0	6	7
19	5	1	0	0	1	3	7
20	5	0	1	1	0	2	5
21	5	0	0	1	0	2	4
22	4	1	0	0	0	2	3
23	2	0	0	1	0	2	4
24	1	0	0	1	0	4	2
25	0	0	0	1	0	2	4
26	0	0	0	0	0	2	3
27	0	1	1	0	0	1	0
28	1	0	1	0	0	1	0
29	2	0	1	0	0	0	0
30	0	0	0	1	0	1	1

Table 8 presents the frequency distribution of consumer purchases by purchase occasion for the actual Sioux Falls data. Note that on purchase occasion 1 there are 1,150 observations and that this number declines with each purchase occasion. This is because all consumers in the data have one purchase occasion, somewhat fewer have two, and so forth.

The upper left cell in Table 9 presents predicted purchase frequencies for the first 10 purchase occasions obtained by simulating data using Model 1—the GL-type model. The simulated sample is the same size as that used in the estimation, and exactly the same exogenous variables that were seen in the dataset were used for the **X** and **A** vectors in the simulation.

In the upper right cell, I use Model 1 to simulate a scenario in which all consumers face a price promotion of 50 cents for alternative 2, Hunt's, on their second purchase occasion. This is a substantial price cut because the average price for the 32-ounce Hunt's during the sample period was roughly $1.10. The GL-type model predicts an immediate increase from 145 to 372 units in Hunt's sales—a 257% increase. It also predicts increases in Hunt's sales in periods 3–10 of 24%, 18%, 13%, 10%, 13%, 14%, 15%, and 15%. (It should be remembered that these numbers include simulation error, so they do not represent exact predictions of sales changes.)

The second row of Table 9 contains the same information for Model 2—the model with κ heterogeneity only, which is quite similar to a familiar Dirichlet multinomial model. This model predicts an immediate increase from 140 to 409 units in Hunt's sales in period 2. Of course, it predicts a return to exactly the previous level of sales as soon as the price promotion is removed—as must any model without state dependence regardless of the complexity of the error structure. The third row presents the simulation results for Model 10—the model that contains all seven types of heterogeneity but does not allow for state dependence. This model predicts an immediate increase from 152 to 409 units in Hunt's sales in period 2. Of course, it also predicts a return to exactly the previous level of sales as soon as the price promotion is removed. Model 10 is quite similar to that of Allenby and Lenk (1994) because it includes AR(1) errors. But, because there is no state dependence, there are no purchase carryover effects by construction.

Finally, the bottom cell of Table 9 presents the simulation results for Model 16—the full model that contains seven types of heterogeneity and true state dependence. The full model predicts that the price cut leads to an immediate increase from 143 to 448 units in Hunt's sales in period 2—a 313% increase. It also predicts increases in Hunt's sales in periods 3–10 of 16%, 12%, 16%, 11%, 5%, 9%, 10%, and 17%. Note that the full model predicts a larger immediate price promotion effect than does the GL-type model. It also predicts a long-run promotion effect on sales that is roughly 40% of the size of the effect predicted by the GL-type model (i.e., over all eight purchase occasions following the promotion, Model 16 predicts a 12% increase in total sales, whereas Model 1 predicts a 17% increase).

Note that the immediate price promotion effect predicted by Model 16 is only a bit larger than that predicted by Model 2—the κ-heterogeneity-only model. More subtly, however, the two models differ dramatically in where they predict the extra sales come from. The full model predicts that 165 out of the 315-unit increase in Hunt's 32-ounce sales come from Heinz 32-ounce (alternative 7), although only 71 units come from the Heinz 28-ounce plastic. The κ-heterogeneity-only model predicts that only 103 units out of a 269-unit increase come from Heinz 32-ounce, but 94 units come from Heinz 28-ounce plastic. The Heinz 40-ounce and 64-ounce are much harder hit according to the κ-heterogeneity-only model, whereas Del Monte 32-ounce and Heinz 32-ounce are much harder hit according to the full model. This occurs because, according to the **L**-vector estimates in the full model, Hunt's 32-ounce is most simi-

Table 9. Price-Policy Experiments

Model	Purchase occasion	Baseline simulation 1	2	3	4	5	6	7	Simulation of 50¢ price cut by Hunt's in period 2: 1	2	3	4	5	6	7
Model 1:	1	191	84	100	41	50	324	360	191	84	100	41	50	324	360
GL form	2	145	62	76	33	41	262	287	372	40	41	18	21	197	217
of state	3	109	40	56	24	19	220	232	135	40	54	20	19	211	221
dependence	4	89	28	39	15	16	156	207	105	27	41	15	17	147	198
but no	5	58	22	27	20	6	143	150	66	20	26	21	7	136	150
heterogeneity	6	51	15	20	8	8	111	116	56	14	18	10	8	111	112
	7	39	5	14	1	3	102	94	44	6	14	2	3	99	90
	8	29	5	10	1	2	61	90	33	5	8	3	2	61	86
	9	20	4	9	3	0	57	66	23	4	9	3	0	58	62
	10	13	3	10	0	4	48	53	15	3	9	0	5	47	52
Model 2:	1	174	66	95	27	42	338	408	174	66	95	27	42	338	408
Kappa-type	2	140	50	74	24	40	268	310	409	31	48	13	24	174	207
heterogeneity	3	111	37	61	17	30	223	221	111	37	61	17	30	223	221
only and no	4	84	34	53	19	19	149	192	84	34	53	19	19	149	192
state	5	57	26	36	20	13	137	137	57	26	36	20	13	137	137
dependence	6	50	13	25	11	20	102	108	50	13	25	11	20	102	108
	7	41	11	23	7	11	86	79	41	11	23	7	11	86	79
	8	35	11	19	8	11	56	58	35	11	19	8	11	56	58
	9	24	9	14	7	8	52	45	24	9	14	7	8	52	45
	10	23	6	12	3	11	39	37	23	6	12	3	11	39	37
Model 10:	1	183	55	95	26	56	350	385	183	55	95	26	56	350	385
Kappa-type	2	152	45	76	24	30	271	308	409	24	54	21	17	210	171
heterogeneity	3	117	35	44	23	30	220	231	117	35	44	23	30	220	231
and random	4	91	33	41	23	17	148	197	91	33	41	23	17	148	197
effects and	5	64	22	36	26	10	118	150	64	22	36	26	10	118	150
AR(1) error	6	64	17	29	16	17	87	99	64	17	29	16	17	87	99
but no state	7	42	8	20	11	6	81	90	42	8	20	11	6	81	90
dependence	8	38	7	18	7	6	53	69	38	7	18	7	6	53	69
	9	32	11	13	6	3	45	49	32	11	13	6	3	45	49
	10	23	5	14	6	5	34	44	23	5	14	6	5	34	44
Model 16:	1	168	64	106	26	60	358	368	168	64	106	26	60	358	368
Full model	2	143	53	71	24	41	274	300	448	22	50	22	26	203	135
	3	113	42	59	20	39	211	216	131	39	58	20	39	210	203
	4	85	37	42	19	20	148	199	95	37	42	17	20	147	192
	5	58	27	38	21	17	121	144	67	28	37	21	17	121	135
	6	62	14	29	13	13	85	113	69	15	30	13	13	84	105
	7	44	15	20	10	6	69	94	46	17	21	10	6	68	90
	8	35	10	17	6	5	51	74	38	10	17	6	5	51	71
	9	29	11	17	7	4	44	47	32	11	17	7	4	44	44
	10	23	6	16	4	3	35	44	27	6	16	5	3	34	40

NOTE: For the four estimated models, simulations were performed to generate unconditional choice sequences. The numbers reported in the table are aggregate choice counts for each of the seven brands on each of the first 10 purchase occasions, based on these simulations.

lar to Heinz 32-ounce and Del Monte 32-ounce, while being rather dissimilar to the other alternatives.

Finally, I note that in the marketing literature several studies have compared the repeat-purchase probabilities of consumers who bought a particular brand on price promotion versus consumers who bought the brand under normal circumstances. It is generally found that repeat purchase rates are lower for consumers who bought on promotion. This finding has often been taken as evidence that buying a brand on promotion actually reduces the probability that a consumer will buy that brand in the future—which would seem to contradict my finding that a temporary price promotion has a small positive effect on future-purchase probabilities. In fact, there is no contradiction. As Neslin and Shoemaker (1989) pointed out, the apparent negative effect of a promotion on repeat-purchase rates may well be due to spurious state dependence that arises out of failure to control for unobserved heterogeneity. The only way to draw valid causal inferences about long-term effects of promotions is to do what I have done in this article: Estimate a choice model that allows for both unobserved heterogeneity and state dependence and then simulate the dynamic consumer response to promotions.

8. CONCLUSION

In this article I have estimated statistical models for consumer choice behavior that allow for both state dependence of the type considered by Guadagni and Little (1983) and a very flexible pattern of heterogeneity. The most general model considered is the first in the brand-choice literature to simultaneously include heterogeneous preferences for common and unique attributes of brands, allow for time-invariant preference heterogeneity and autocorrelated time-varying preference shocks, and include state dependence (or purchase-event feedback). Using the Nielsen ketchup data, I find that the parameters capturing state dependence remain highly significant even after controls for very complex patterns of heterogeneity. This result lends support to the notion that choice is not a zero-order process.

But the results also indicate that in a certain sense the deviation of choice behavior from a zero-order process is quantitatively small. Simulation results for the preferred model reveal that a substantial price reduction large enough to generate a 313% immediate sales increase in one of the ketchup brands leads to only about a 12% sales increase on subsequent purchase occasions after price returns to normal. The increment to a consumer's current utility evaluation for a brand of ketchup due to having bought that brand on the previous purchase occasion is roughly equal to what would be achieved by a 5-cent price cut—which is small relative to the variability of prices in the data. Thus, the effects of current prices on current utility evaluations appear to be much stronger than purchase carryover effects. This relative weakness of purchase carryover effects probably accounts for the fact that most previous research in marketing has failed to reject the zero-order hypothesis (see Bass 1993). Given the weakness of purchase carryover effects relative to current-price effects and other marketing-mix effects, highly efficient estimation techniques like those employed here are needed to detect state dependence in the data.

ACKNOWLEDGMENTS

I thank seminar participants at Wharton, MIT, NYU, Rochester, Berkeley, and Stanford for comments on earlier versions of this work. The data used in this study were provided by A. C. Nielsen Corporation. The Federal Reserve Bank of Minneapolis provided computer support. The views expressed herein are not necessarily those of the Federal Reserve System.

[Received January 1995. Revised March 1996.]

REFERENCES

Allenby, G., and Lenk, P. (1994), "Modeling Household Purchase Behavior With Logistic Normal Regression," *Journal of the American Statistical Association*, 89, 1218–1231.

Avery, R., Hansen, L., and Hotz, V. J. (1983), "Multiperiod Probit Models and Orthogonality Condition Estimation," *International Economic Review*, 24, 21–35.

Bass, F. (1993), "The Future of Research in Marketing: Marketing Science," *Journal of Marketing Research*, 30, 1–6.

Bass, F., Jeuland, A., and Wright, G. (1976), "Equilibrium Stochastic Choice and Market Penetration Theories: Derivations and Comparisons," *Management Science*, 22, 1051–1063.

Chintagunta, P. (1994), "Heterogeneous Logit Model Implications for Brand Positioning," *Journal of Marketing Research*, 31, 304–311.

Elrod, T. (1988), "Choice Map: Inferring a Product Market Map From Panel Data," *Marketing Science*, 7, 21–40.

Elrod, T., and Keane, M. (1995), "A Factor Analytic Probit Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32, 1–16.

Erdem, T. (1996), "A Dynamic Analysis of Market Structure Based on Panel Data," *Marketing Science*, 15, 359–378.

Erdem, T., and Keane, M. (1996), "Decision Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15, 1–20.

Fader, P. (1993), "Integrating the Dirichlet-multinomial and Logit Models of Brand Choice," *Marketing Letters*, 4, 99–112.

Fader, P., and Lattin, J. (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-sectional Model of Consumer Purchase Behavior," *Marketing Science*, 11, 372–385.

Geweke, J. (1991), "Efficient Simulation From Multivariate Normal and Student-*t* Distributions Subject to Linear Constraints," in *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, Alexandria, VA: American Statistical Association, pp. 571–578.

Geweke, J., Keane, M., and Runkle, D. (1994), "Alternative Computational Approaches to Statistical Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, 76, 609–632.

——— (in press), "Statistical Inference in the Multinomial Multiperiod Probit Model," *Journal of Econometrics*, 1997.

Guadagni, P., and Little, J. (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, 203–238.

Hajivassiliou, V., McFadden, D., and Ruud, P. (1994), "Simulation of Multivariate Normal Orthant Probabilities: Methods and Programs," DP #1021, Yale University, Cowles Foundation.

Heckman, J. (1981a), "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. S. Rosen, Chicago: University of Chicago Press, pp. 91–139.

——— (1981b), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data With Economic Applications*, eds. C. Manski and D. McFadden, Cambridge, MA: MIT

Press, pp. 179–197.

Jones, J., and Landwehr, T. (1988), "Removing Heterogeneity Bias From Logit Model Estimation," *Marketing Science*, 7, 41–59.

Kamakura, W., and Russell, G. (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26, 379–390.

Keane, M. (1990), "Four Essays in Empirical Macro and Labor Economics," unpublished Ph.D. dissertation, Brown University, Dept. of Economics.

——— (1992), "A Note on Identification in the Multinomial Probit Model," *Journal of Business & Economic Statistics*, 10, 193–200.

——— (1993), "Simulation Estimation for Panel Data Models With Limited Dependent Variables," in *The Handbook of Statistics 11: Econometrics*, eds. C. R. Rao, G. S. Maddala, and H. D. Vinod, Amsterdam: Elsevier, pp. 545–572.

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

McCulloch, R., and Rossi, P. (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 64, 207–240.

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

Neslin, S., and Shoemaker, R. (1989), "An Alternative Explanation for Lower Repeat Rates After Promotion Purchases," *Journal of Marketing Research*, 26, 205–213.

Steckel, J., and Vanhonacker, W. (1988), "A Heterogeneous Conditional Logit Model of Choice," *Journal of Business & Economic Statistics*, 6, 391–398.