Quant Market Econ (2006) 4: 57–81
DOI 10.1007/s11129-006-6559-x

Modeling preference evolution in discrete choice models: A Bayesian state-space approach

Mohamed Lachaab · Asim Ansari · Kamel Jedidi · Abdelwahed Trabelsi

Abstract We develop discrete choice models that account for parameter driven preference dynamics. Choice model parameters may change over time because of shifting market conditions or due to changes in attribute levels over time or because of consumer learning. In this paper we show how such preference evolution can be modeled using hierarchial Bayesian state space models of discrete choice. The main feature of our approach is that it allows for the simultaneous incorporation of multiple sources of preference and choice dynamics. We show how the state space approach can include state dependence, unobserved heterogeneity, and more importantly, temporal variability in preferences using a correlated sequence of population distributions. The proposed model is very general and nests commonly used choice models in the literature as special cases.

We use Markov chain monte carlo methods for estimating model parameters and apply our methodology to a scanner data set containing household brand choices over an eight-year period. Our analysis indicates that preferences exhibit significant variation over the time-span of the data and that incorporating time-variation in parameters is crucial for appropriate inferences regarding the magnitude and evolution of choice elasticities. We also find that models that ignore time variation in parameters can yield misleading inferences about the impact of causal variables.

Keywords Preference evolution · Hierarchical Bayesian state-space models · Heterogeneity · Multinomial probit · Choice models · Pricing · Promotions

This paper is based on the first author's doctoral dissertation.

M. Lachaab
Institut Superieur de Gestion, University of Tunis, Tunisia. Currently: Visiting Faculty, Clarkson University, New York, USA

A. Ansari and K. Jedidi
Graduate School of Business, Columbia University, Uris Hall, 3022 Broadway, New York, NY, USA
e-mail: maa48@columbia.edu

A. Trabelsi
Institut Superieur de Gestion, University of Tunis, Tunisia

1. Introduction

Marketers are often interested in understanding how consumer preferences and choices vary over time. A significant number of studies in the discrete choice literature within marketing and economics have focused on modeling such dynamics. Previous researchers have used a variety of approaches to model preference variation. For instance, some researchers have incorporated dynamics in choice models by allowing alternative-specific utility errors to be serially correlated over time (Allenby and Lenk, 1995; Geweke et al., 1997). Other researchers have modeled state dependence arising due to inertia, habit persistence and carryover effects to capture the linkages between adjacent choice observations (Heckman, 1981; Gaduagni and Little, 1983; Allenby and Lenk, 1994; Roy et al., 1996; Keane, 1997; Seetharaman, 2003). Others have developed discrete choice models in which preference parameters are allowed to vary systematically as a function of past marketing actions based on distributed lag specifications (Mela et al., 1997; Jedidi et al., 1999; Seetharaman, 2003).

In this paper, we contribute to the discrete choice literature by showing how a Bayesian state space framework (West and Harrison, 1997) can be used to augment the above approaches by flexibly incorporating parameter-driven preference dynamics in choice models. In a state space model, the preference parameters are directly allowed to stochastically vary over calendar time in a structured fashion. This allows the researcher to understand how parameters vary over the long-term in an economically meaningful way. State space models induce markov dependence among parameters in successive time periods and thereby capture the evolution of preferences over time. By allowing the parameters to stochastically as well as systematically vary over time, the state-space approach captures both observed and unobserved sources of temporal variation. The unobserved sources represent the impact of temporally varying random shocks experienced in different time periods. Including such unobserved sources of parameter variation provides an important advantage over models that allow parameters to evolve solely as a function of previous marketing actions.

The state space approach, moreover, does not preclude the inclusion of other forms of dynamics. For instance, state dependence specifications that have been used in the past can be easily incorporated as part of the state space approach that we outline in this paper. State dependence models typically use lagged dummy variables or exponentially smoothed loyalty specifications to capture the impact of previous choices on current utilities and to specify the linkages between successive choice observations of the same individual. In addition to capturing state dependence, the state space model incorporates the dynamics occurring at a different time scale (e.g., weeks, months). Specifically, the model parameters (including state dependence parameters) can vary over calendar time thus facilitating a clearer understanding of the macro evolution of tastes and preferences in the market place.

Choice model parameters can differ over calendar time because of a variety of reasons. For example, market conditions may vary over the long-run, and thus market preferences may change due to shifting conditions. Changing economic conditions can result in consumers becoming more or less price and promotion sensitive over time. Specific brands may go out of favor and others may gain in popularity because of contagion and word of mouth effects. It is also possible that attribute levels available in the market-place may change over time because of managerial actions. The behavioral literature on decision-making indicates that the perceived importance of an attribute depends upon its range. Thus, variations in attribute ranges can results in changes in coefficients and in the utilities for choice alternatives. Changes in parameters may also result because of consumer learning (Erdem, 1996) and shaping of preferences. For example, the literature on the long-term effects of advertising and promotions

(Jedidi et al., 1999) suggests that in the long-run, increased exposures to price promotions can make consumers more price sensitive in the market-place.

Understanding how parameters vary over time is critical for optimizing marketing decisions. For instance, an increase in consumers' price sensitivity over time suggests decreasing optimal price levels. When preferences vary over time, but the model assumes a constant set of parameters for the entire time-span of the data, an important source of variability is left unspecified. It is well understood that ignoring cross-sectional heterogeneity results in misleading inferences about consumer preferences. In this paper, we show that modeling the dynamic variation of preferences is also crucial. Ignoring such dynamics can result in misleading inferences regarding the temporal pattern of elasticities. Such erroneous inferences can translate into sub-optimal marketing over time.

In this paper, we develop and empirically test a general state space, heterogeneous, multinomial probit model (MNP). We show how a variety of MNP models can be obtained as special cases of our model. Methodologically, our paper contributes to the growing Bayesian estimation literature on discrete choice models (McCulloch and Rossi, 1994; McCulloch et al., 2000; McCulloch and Rossi, 2000). We show how different forms of temporal variation and cross-sectional heterogeneity can be simultaneously incorporated in the multinomial probit model via a dynamic population distribution. We also show how Bayesian inference can be conducted for our model and illustrate how a combination of data augmentation (Albert and Chib, 1993), Gibbs sampling, Metropolis-Hastings and Slice sampling (Neal, 2003) methods can be used for MCMC estimation of the model parameters.

Substantively, we investigate preference evolution using an eight year panel data involving household purchases of a packaged good category. We also compare our results to those obtained from several null models. The results show that models that jointly account for both consumer heterogeneity and temporal dependence statistically outperform models that ignore any one of these phenomena. In addition, we show how consumer sensitivities to the marketing variables evolve during the eight year time span. Our results yield interesting patterns of variation over time in the brand intercepts and marketing-mix parameters. Finally, our results illustrate how models that ignore dynamics can yield misleading inferences regarding the evolution of preferences.

The rest of the paper is organized as follows. In the next section, we present the heterogenous state-space choice model. In Section 3, we discuss Bayesian inference for the proposed model. In Section 4 we describe the application of our model to scanner panel data. We show how the model can be used to understand the temporal pattern of elasticities and discuss the managerial implications of the results. Finally, in Section 5 we conclude the paper by outlining the limitations inherent in our work and offer suggestions for future research.

2. The hierarchical state space model

2.1. Multinomial probit

We begin with a multinomial probit specification. Consider a panel of H households choosing one brand among a set of $M + 1$ mutually exclusive brands on each purchase occasion. In many marketing data sets such as those obtained from supermarket scanner panels, choice information is available over multiple time periods. We use $t = 1$ to T to index calendar time (e.g., weeks or months). Let y_{hj} indicate the observed multinomial discrete choice on the jth observation of household h. If household h contributes n_h such observations (i.e., $j = 1$ to n_h), the total number of observations in the data are given by $N = \sum_{h=1}^{H} n_h$. The discrete choices can be modeled using a random utility framework (McFadden, 1977). Let $\mathbf{u}^*_{hj} = \{u^*_{hj1}, u^*_{hj2}, \ldots, u^*_{hjM+1}\}$ denote the vector of brand-specific utilities underlying the choices. Then, $y_{hj} = m$, if brand m has the highest utility. Consistent with a random utility formulation, the utility for a brand m can be expressed in terms of a systematic component composed of brand-specific variables contained in a vector $\mathbf{w}_{hjm}$ and a stochastic component representing the impact of unobserved influences.

As is well known, because the observed responses are multinomial, only differences in utilities are identified from the data. We therefore represent the responses in terms of a $M \times 1$ vector $\boldsymbol{u}_{hj} = \{u_{hj1}, \ldots, u_{hjM}\}$ which contains differenced utilities with respect to a base alternative. Specifically, the elements, $u_{hjm} = u^*_{hjm} - u^*_{hjM+1}$ denote differences with respect to an arbitrary chosen brand, for example, the $(M+1)$th brand. The differenced utility equations can be written as

$$u_{hjm} = \mathbf{x}'_{hjm}\boldsymbol{\beta}_{hj} + e_{hjm}, \tag{1}$$

where, e_{hjm} is an appropriately differenced error term. The vector $\mathbf{x}_{hjm}$ contains M brand-specific dummy variables for the brand intercepts and the differenced brand-specific variables $w_{hjm} - w_{hj\mathrm{M}+1}$ for the mth brand. The resulting M equations can be stacked together in matrix form to yield a system of equations

$$\mathbf{u}_{hj} = \mathbf{X}_{hj}\boldsymbol{\beta}_{hj} + \mathbf{e}_{hj}, \tag{2}$$

where, $\mathbf{u}_{hj}$ is a $M \times 1$ vector of utilities, and $\mathbf{X}_{hj}$ is a $M \times L$ matrix appropriately composed from the vectors $\mathbf{x}_{hjm}$. As noted before, "observed" dynamics can be included by using explanatory variables that capture state dependence (e.g., lagged dummy indicators, loyalty, etc). These become part of the $\mathbf{X}_{hj}$ matrix in Eq. (2) and are used to specify the linkages between successive choice observations (i.e., observations j and $j+1$) for the same household. The $L \times 1$ coefficient vector $\boldsymbol{\beta}_{hj}$ contains the M brand-specific intercepts which represent the intrinsic preferences for the brands (relative to the $(M+1)$th brand) and the response coefficients for the explanatory variables.

The stochastic components contained in the error vector $\mathbf{e}_{hj}$, represent the combined influence of myriad factors that are known to the consumer but are unobserved to the researcher. As the utilities are in differenced form, and because unobserved factors may be common across brands, we assume that the errors $\mathbf{e}_{hj}$ are distributed multivariate normal $N(\mathbf{0}, \boldsymbol{\Sigma})$, where, the variance matrix $\boldsymbol{\Sigma}$ captures the covariation in the errors across the equations. The differencing of utilities takes care of the locational indeterminacies of the underlying utilities. Model identification also requires that we fix the scale of the utilities. This can be achieved by setting one of the variances in $\boldsymbol{\Sigma}$ to one. We now show how the state space approach can be incorporated in the above described standard multinomial probit specification.

2.2. Population distribution

When choice data are available for multiple time periods $(t = 1, \ldots, T)$, it is imperative to explicitly model the evolution of preferences over the time span of the data. To model such dynamics, we assume that the multinomial probit parameters $\boldsymbol{\beta}_{hj}$ are drawn from a population distribution that evolves from one period to another. A temporally changing population distribution is ideal for capturing the dynamics in population preferences as it enables a proper accounting of the different sources of variation in choices across the observations.

To formulate this population distribution, we assume that the coefficient vector $\boldsymbol{\beta}_{hj}$ can be decomposed into two parts according to the equation

$$\boldsymbol{\beta}_{hj} = \boldsymbol{\mu}_{t_{hj}} + \mathbf{b}_h. \tag{3}$$

In Eq. (3), the first part $\boldsymbol{\mu}_{t_{hj}}$ represents the mean of the temporally varying population distribution. The population mean is indexed by the subscript t_{hj} which identifies the calendar time period pertaining to the jth observation of household h. The subscript t_{hj} takes values in the set $\{1, 2, \ldots, T\}$. Notice that observations for household h are subscripted with (hj) and calendar time periods are subscripted with t_{hj}. This distinction is necessary to differentiate between two different time scales and to establish a correspondence between observations in the data and the associated time periods. Because of the unbalanced nature of the data across households, not every household is observed to purchase in every time period. Thus, the jth observation of household h may occur in time period t, i.e., $t_{hj} = t$, whereas, the jth observation of another household h' may occur in time period t', i.e., $t_{h'j} = t'$. Notice that, if every consumer were to be observed in every time period only once, then the two time scales would be equivalent. This is, however, not the case in many data sets (such as those obtained from supermarket scanner panels), and therefore, calendar time offers a useful mechanism to coalesce such disparate data across different consumers.

The second component of $\boldsymbol{\beta}_{hj}$ in Eq. (3), is a household-specific vector $\mathbf{b}_h$ that contains all the unobserved influences specific to household h that impact its preferences. Notice that this component is invariant across time periods and can be used to characterize the cross-sectional variation in response coefficients across households. As usual, we assume that the random effects $\mathbf{b}_h$ are distributed multivariate normal $\mathbf{b}_h \sim N(\mathbf{0}, \boldsymbol{\Omega})$. The covariance matrix $\boldsymbol{\Omega}$ is assumed to be common across all households and is time-invariant. Note that $\mathbf{b}_h$ has a zero mean. This assumption is necessary for identification as the household-specific parameters, $\mathbf{b}_h$, are specified as deviations from the time varying population mean $\boldsymbol{\mu}_{t_{hj}}$.

Together, these two components imply that the response parameters at any observation j for household h are drawn from a temporally varying population distribution

$$\boldsymbol{\beta}_{hj} \sim N\left(\boldsymbol{\mu}_{t_{hj}}, \boldsymbol{\Omega}\right). \tag{4}$$

Given T time periods in the data, there are T such population distributions, one associated with each time period. Until now, our description pertained to a standard hierarchical probit model, with the exception that we allow the population distribution to differ across time periods. We now show how these population distributions are related.

2.3. Population dynamics

The T population distributions, together, specify the population dynamics. However, for capturing the incremental evolution of preferences from one time period to another, we need to establish a link between population distributions from adjacent time periods. We do so by assuming that the population mean in any time period is stochastically related to the population mean in the previous time period. Specifically, we posit a general vector autoregressive process to characterize the dynamics of the population mean. This implies, that the law of motion for the population mean can be written according to the transition equation[1]:

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \mathbf{Z}_t\boldsymbol{\theta} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Psi}), \tag{5}$$

where, $\boldsymbol{\mu}_t$ is the population mean for an arbitrary time period t, $\boldsymbol{\mu}_{t-1}$ is the population mean vector from the previous time period, and $\boldsymbol{\Phi}$ represents the transition matrix. In addition to modeling the relationship between the population means for successive time periods, the above VARX(1) specification also captures the systematic impact of variables on the transition process via the term $\mathbf{Z}_t\boldsymbol{\theta}$. The matrix $\mathbf{Z}_t$ contains constants for the intercepts and explanatory variables that influence the evolution. When $\mathbf{Z}_t$ is an identity matrix, $\boldsymbol{\theta}$ contains only the intercepts in the vector auto-regression. The transition process is assumed to start with an initial state $\boldsymbol{\mu}_0$ that is distributed $N(\boldsymbol{\eta}_0, \mathbf{C}_0)$. Finally, the vector $\boldsymbol{\epsilon}_t$ contains unobserved variables that represent temporally varying random shocks experienced in that time period t. We assume that $\boldsymbol{\epsilon}_t$ is independent of all utility errors $\mathbf{e}_{hj}$ and the cross-sectional heterogeneity vector $\mathbf{b}_h$. The covariance matrix $\boldsymbol{\Psi}$ for the transition errors is of dimension $L \times L$ and is assumed to be constant over time. The different elements of $\boldsymbol{\Psi}$ delineate how different pairs of elements in $\boldsymbol{\mu}_t$ co-vary. The dynamics in Eq. (5), when combined with our assumptions about population heterogeneity, in Eq. (4), imply that coefficients for different individuals evolve in the same manner as the population mean. This assumption may not be realistic in certain situations, but allows us to parsimoniously structure the dynamics, especially considering that in our data, not every individual makes a purchase in each time period.

From Eq. (5), we see that upon repeated substitution, the population mean $\boldsymbol{\mu}_t$ can be written as a weighted sum of the initial state $\boldsymbol{\mu}_0$, all the previous exogenous variables in $\mathbf{Z}_t$, and all the past transition errors $\boldsymbol{\epsilon}_t$ as follows:

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}^t\boldsymbol{\mu}_0 + \sum_{l=0}^{t-1}\boldsymbol{\Phi}^l\mathbf{Z}_{t-l}\boldsymbol{\theta} + \sum_{l=0}^{t-1}\boldsymbol{\Phi}^l\boldsymbol{\epsilon}_{t-l}, \tag{6}$$

where $\boldsymbol{\Phi}^0 = \mathbf{I}$. Then the expectation of $\boldsymbol{\mu}_t$ is given by

$$\mathrm{E}(\boldsymbol{\mu}_t) = \boldsymbol{\Phi}^t\boldsymbol{\eta}_0 + \sum_{l=0}^{t-1}\boldsymbol{\Phi}^l\mathbf{Z}_{t-l}\boldsymbol{\theta}, \tag{7}$$

and its variance can be written as

$$\mathrm{Var}(\boldsymbol{\mu}_t) = \boldsymbol{\Phi}^t\mathbf{C}_0(\boldsymbol{\Phi}^t)' + \sum_{l=0}^{t-1}\boldsymbol{\Phi}^l\boldsymbol{\Psi}(\boldsymbol{\Phi}^l)'. \tag{8}$$

Many different forms of the dynamics can be obtained from Eq. (5) as special cases. We now describe briefly these special cases.

[1] Notice that in this section, we focus on the population distribution. Hence, subscripts pertaining to particular consumers are not relevant.

Dynamic random effects

Consider the situation, where $\mathbf{\Phi} = \mathbf{0}$, and $\mathbf{Z}_t = \mathbf{I}$. In this case, we see that

$$\boldsymbol{\mu}_t = \boldsymbol{\theta} + \boldsymbol{\epsilon}_t. \tag{9}$$

This yields a dynamic system wherein, the population distribution is assumed to evolve independently from one time period to another, and the expectation and variance of $\boldsymbol{\mu}_t$ are constant across time periods. The process assumes zero memory and random shocks are absorbed within the same period.

Random walk

When $\mathbf{\Phi} = \mathbf{I}$, and $\boldsymbol{\theta}$ is zero, we obtain the random walk model. The transition equation has the form

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t, \tag{10}$$

i.e., the new population mean is obtained from the previous mean with the addition of some white noise. Notice that shocks do not attenuate over time, and from Eq. (8), we see that $\mathrm{V}(\boldsymbol{\mu}_t) = \mathbf{C}_0 + t\Psi$. Thus the model assumes that the variance of the coefficients increases directly with t. However, the conditional variance $\mathrm{V}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1})$ is constant at $\mathbf{\Psi}$. The above disadvantages are offset by the fact that the random walk model allows for a smooth evolution of the parameters, has a simple interpretation and yields a parsimonious specification for capturing the dynamics in the population mean.

Vector autoregression

When $\mathbf{Z}_t = \mathbf{I}$, we obtain the vector autoregressive model VAR(1) of order one. The influence of the initial condition fades away if the largest eigenvalue of $\mathbf{\Phi}$ is less than unity in magnitude. In this instance, the transition process can be considered asymptotically stationary (Lutkepohl, 1991, c. 11). Restricted versions of the autoregressive model can be obtained by allowing $\mathbf{\Phi}$ to be diagonal. In this case, a coefficient at time t is assumed to be related only to its own value at time $t-1$, and this yields a parsimonious specification with fewer parameters to estimate.

To summarize, when population dynamics are combined with cross-sectional heterogeneity, the resulting general model can be presented as follows

$$\mathbf{u}_{hj} = \mathbf{X}_{hj}\boldsymbol{\beta}_{hj} + \mathbf{e}_{hj}, \tag{11}$$

$$\boldsymbol{\beta}_{hj} = \boldsymbol{\mu}_{t_{hj}} + \mathbf{b}_h, \tag{12}$$

$$\mathbf{e}_{hj} \sim N(\mathbf{0}, \mathbf{\Sigma}), \tag{13}$$

$$\mathbf{b}_h \sim N(\mathbf{0}, \mathbf{\Omega}), \tag{14}$$

$$\boldsymbol{\mu}_t = \mathbf{\Phi}\boldsymbol{\mu}_{t-1} + \mathbf{Z}_t\boldsymbol{\theta} + \boldsymbol{\epsilon}_t, \tag{15}$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{\Psi}), \tag{16}$$

where, the errors $\mathbf{e}_{hj}$, ϵ_t, and $\mathbf{b}_h$ are assumed to be mutually independent. This results in a state-space, hierarchical multinomial probit model which is built up from a series of conditional distributions that combine to yield a multi-level specification. At the bottom level of the hierarchy is the latent regression model conditional on β_{hj} and $\mathbf{\Sigma}$ (see Eq. (11)). The random utility regression is followed by successively higher levels of priors which incorporate views about the distributions of $\boldsymbol{\mu}_t$ and $\mathbf{b}_h$. The state space form of the model (Priestly, 1980; West and Harrison, 1997) is described by Eqs. (11) and (15). The former equation represents the "observation" equation of the dynamic system. Given that the utilities are latent, this is observed only in a data-augmentation sense. The latter equation is called the state (transition) equation. It describes the markov dependence of the states $\boldsymbol{\mu}_t$ over time periods.

The hierarchical state space probit model is very general and subsumes several well-known choice models as special cases. If $\mathbf{\Phi} = \mathbf{0}$, $\mathbf{Z}_t = \mathbf{I}$, $\mathbf{\Psi} = \mathbf{0}$ and $\mathbf{\Omega} = \mathbf{0}$, then the probit coefficients do not vary across households or time periods, and the model reduces to the traditional static multinomial probit model (Hausman and Wise, 1978). If $\mathbf{\Omega} = \mathbf{0}$, and no other restrictions are employed, then a state space (dynamic) model without consumer heterogeneity is obtained. This represents a situation where household heterogeneity is negligible, but the fixed effect parameters are assumed to be dynamic. This situation is the one most often considered in the state-space literature, as most applications of the state space model have time series data on a single cross-section. A static heterogeneous multinomial probit model without dynamics results if $\mathbf{\Phi} = \mathbf{0}$, $\mathbf{Z}_t = \mathbf{I}$, $\mathbf{\Psi} = \mathbf{0}$, but $\mathbf{\Omega}$ is unrestricted (McCulloch and Rossi, 1994) which is the defacto standard in the marketing literature on discrete choice models.

3. Bayesian inference

The likelihood for the hierarchical state space probit model is exceedingly complex as it involves multiple integration of a very high dimensionality. The complexity arises from many different sources. First, the probabilities associated with the multinomial probit require integrations involving multivariate normal CDF's. Second, the random effects need to be integrated out. This requires incorporating the multiple population distributions. Moreover, this needs to be done, taking into consideration the time series relationships between the population means. Given these sources of complexity, it is virtually impossible to use classical methods for estimation.

Bayesian methods, however, lead to considerable simplifications. These occur because of the possibility of data augmentation in simulation based Bayesian inference. Once utilities are "known", the model becomes linear in the preference parameters. Thus, conditional on the utilities and the random effects, the state parameters (i.e., the population means) can be obtained based on the Kalman filter (Kalman, 1960). Given the hierarchical nature of the model, Bayesian methods are a natural alternative as they allow for proper accounting of all uncertainties in the estimation of the random effects and the state parameters.

We use Markov chain Monte Carlo methods (MCMC) involving a combination of Gibbs sampling (Gelfand and Smith, 1990), Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970), slice sampling (Neal, 2003) and data augmentation (Albert and Chib, 1993; Tanner and Wong, 1987) steps to obtain the requisite sample of parameter draws from the joint posterior distribution. In using MCMC methods, the choice of appropriate blocking strategies is important for adequate mixing of the resulting Markov chain. This is specially critical in sampling the state vectors (population means). If each of the state vectors in $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots \boldsymbol{\mu}_T\}$

are simulated from its full conditional distribution one at a time, convergence to the target distribution could be very slow, especially if the dimension T is large. To circumvent this problem, we use the forward-filtering backward-sampling algorithm (Fruhwirth-Schnatter, 1994 and Carter and Kohn, 1994) to jointly simulate all the states in one block from the joint full conditional distribution of the states. To sample the covariance matrix of the utilities, we follow Barnard et al. (2000) and use a combination of slice sampling and Metropolis-Hastings steps for obtaining the correlations and the log-standard deviations, respectively. Slice sampling is particularly suitable for the correlations because one can then easily restrict the correlations to an interval that respects positive-definiteness of the covariance matrix. Moreover, it does not require any tuning, thus providing an automatic method for obtaining the correlations. The mathematical details for the priors and the full conditional distributions are described in the Appendix.

4. Application

4.1. Data and variables

We use a scanner panel dataset that is collected in a medium sized mid-western market in the United States. The category can be described as a frequently purchased, household, non-food good.[2] The data contain choices for four brands from eleven stores over a duration of eight years from 1984 to 1992. The demographics of the panel do not deviate substantially from the national averages. The panelists constitute a static sample where households did not enter or exit the panel during the period of analysis. We have data on 300 households who made 5781 purchases in total for the four brands over 91 months. The households' mean interpurchase time for the product category is about 12 weeks. In addition, no brand entries or exits occurred over the duration of the data. Consequently, product introduction factors and product life-cycle issues are not likely to impact our analysis. The mean number of observations per household is 19 with a standard deviation of 14 whereas, the mean number of observations per month is 63 with a standard deviation of 14.[3]

In addition to household purchase data, information on price and price promotion of the four brands are available in the data set. The price represents the regular (non-promoted) price, whereas, the price promotion variable reflects the monetary discount offered by a brand during a promotion. Figure 1 shows how market shares, prices, and promotions for the four brands varied over the time span of the data. The figure shows that, except for short-term fluctuations, the market shares for the brands are stable over the long-run. In contrast, the brand prices and promotions show an increasing trend over time with the promotion increase occurring mostly after month eighty. Across brands and time periods, the average price is $0.05 per ounce of the product, the average promotion frequency is 25%, and the average discount is 17% of the regular price.

[2] We cannot reveal the product category and the brand names because of confidentiality agreement with the data provider.

[3] The high mean interpurchase time resulted in a very low number of observations per week. We therefore analyze our data at the monthly level to ensure that we have sufficient number of observations per period for reliably estimating the time varying parameters.

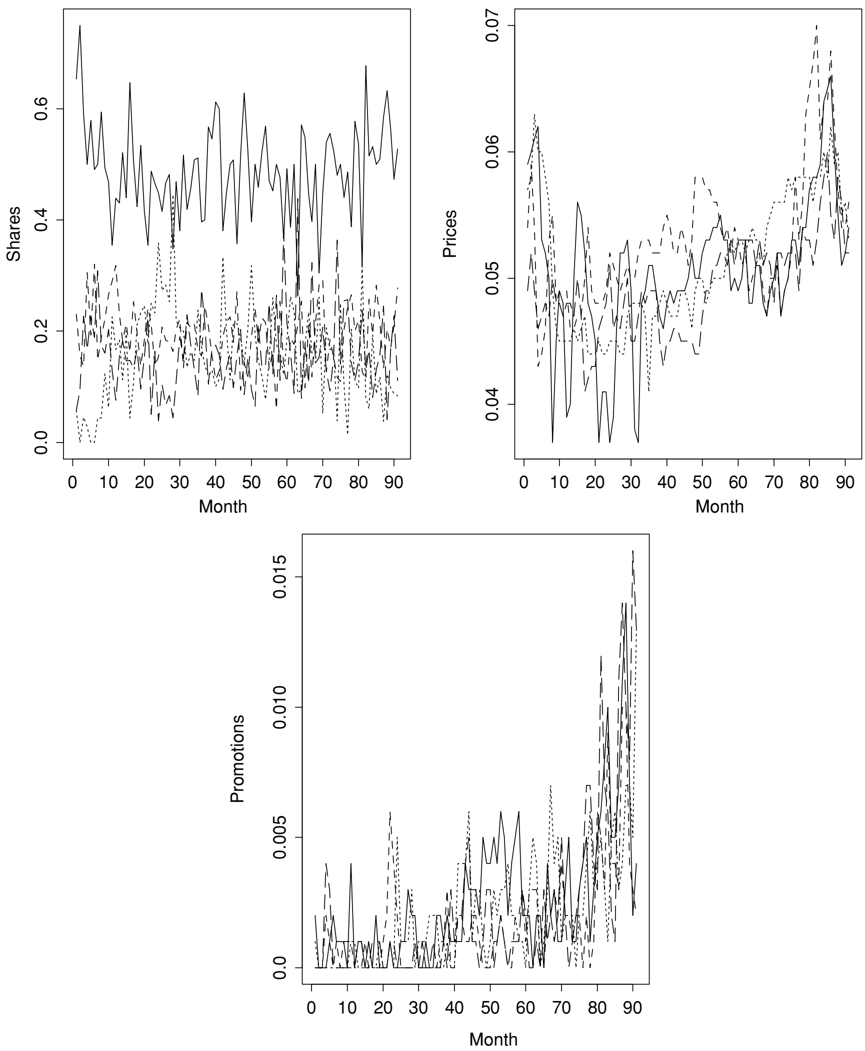

Fig. 1 Time variation in Brand Shares, Prices and Promotions for the four brands over the span of the data. Brand 1 (solid line), Brand 2 (dotted line), Brand 3 (small-dash line) and Brand 4 (long-dash line) models. The price and promotion variables are represented in $/ounce

4.2. Model specifications

We estimate the general, state-space, heterogeneous probit model (Cprobit; see Eqs. (11)–(16)) and three null models using this data set. The null models were selected to understand the effects of parameter time variation and cross-sectional household heterogeneity separately and in tandem. The first null model (Sprobit) is a simple probit model which does not account for either unobserved heterogeneity or time-varying parameters. This forms the base model from which the effects of adding heterogeneity and time-varying parameters can be assessed. In the second null model (Hprobit), we add unobserved sources of heterogeneity to the Sprobit model, but do not allow the population mean to vary over time. Thus, Hprobit is the standard hierarchical probit model that is routinely used in the marketing literature. In the third null model (Tprobit), we add time-varying effects to Sprobit as specified in Eq. (5), but do not include unobserved sources of customer heterogeneity. This yields a (non-heterogeneous) state-space probit model, which assumes that the parameters of the utility function vary over time.

To examine which of the dynamic structures govern the motion of the preference parameters, we estimate three additional special cases of Cprobit. The first (Cprobit-RE) is a heterogeneous, dynamic random-effects model where $\mathbf{\Phi} = \mathbf{0}$, and $\mathbf{Z}_t = \mathbf{I}$ (see Eq. (9)). The second (Cprobit-RW) is a heterogeneous, random walk model (see Eq. (10)). The last model (Cprobit-Diag) is the same as Cprobit but with $\mathbf{\Phi}$ constrained to be diagonal.

In each of the seven models, we use five regressors to specify the systematic component of the utilities for the brands: the two marketing variables: PRICE and PROM, which are the regular price and promotional discount respectively, both measured in cents per ounce, and the three dummy variables to capture the intrinsic preferences of each brand relative to the base alternative. We chose the fourth brand to be the base alternative. Given that we are using a data set with 300 households observed over 91 months, the unknown utilities and parameters for the models are the following: $\{\{\mathbf{u}_{hj}\}\}$ (5781 × 3), $\{\boldsymbol{\mu}_t\}$ (91 × 5), $\{\mathbf{b}_h\}$ (300 × 5), $\mathbf{\Psi}$ (5 × 5), $\mathbf{\Omega}$ (5 × 5), $\mathbf{\Sigma}$ (3 × 3), $\mathbf{\Phi}$ (5 × 5) and $\boldsymbol{\theta}$ (5 × 1). Recall that the first variance in $\mathbf{\Sigma}$, σ_{11}, is fixed to 1 for identification purpose. Depending upon what is included in a model, some of the above mentioned unknowns may not be present as part of its specification. For instance, $\boldsymbol{\mu}_t$ and $\mathbf{\Psi}$ are not part of the Sprobit and Hprobit models.

4.3. Results

In this section, we present the results from the different models. We also discuss how consumer preferences vary over time and across households in our dataset. As explained in Section 3, we used Markov chain monte carlo (MCMC) methods for estimating the seven models. For each model, we ran sampling chains for 150,000 iterations. In each case, convergence was assessed by monitoring the time-series of the draws. The results reported in the paper are based on 100,000 draws retained after discarding the initial 50,000 draws as burn-in iterations. We now report the model comparison results.

4.3.1. Model comparisons

Model comparisons within the Bayesian approach are traditionally done using the Bayes factors (Kass and Raftery, 1995). Bayes factors account for model fit, but also automatically penalize for model complexity. The Bayes factor for comparing two models can be written in terms of the ratio of the observed marginal densities for the models. A simulation consistent estimate for the marginal densities can be obtained using the MCMC draws that are generated for parameter inference. We used these draws to generate an importance sampling estimate of the log-marginal likelihood (Newton and Raftery, 1994) for each of the seven models. Table 1 reports these log-marginal likelihoods.

A comparison of the values in Table 1 sheds light on the need for capturing both unobserved heterogeneity and dynamics. The Cprobit-RW model, which includes both unobserved heterogeneity and time-varying effects (specified as a random walk) has the highest log-marginal likelihood. The largest improvement in log-marginal likelihoods occur when unobserved heterogeneity is added. For instance, Hprobit provides a vast improvement over models that do not include heterogeneity. Improvements due to the addition of dynamics are also significant when one compares the value for Cprobit-RW with that for the Hprobit model. However,

Table 1 Model comparisons

Probit model	Label	Log-marginal likelihood
Simple	Sprobit	−7314.27
Heterogeneous	Hprobit	−4839.31
Dynamic	Tprobit	−6634.63
Full Model	Cprobit	−4560.39
Full with Diagonal Φ	Cprobit-Diag	−4606.84
Full–Random Effects	Cprobit-RE	−4612.84
Full–Random Walk	Cprobit-RW	−4528.91

the variability across consumers appears to be larger than that across time periods. Although the random walk is selected to be the best dynamics structure, the log-marginal likelihood value of Cprobit-RW is relatively close to that of the full model (Cprobit) and does not differ much from those of the others dynamics forms. Thus the data appear to offer a strong signal about the importance of accounting of dynamics but a weaker signal as to the form of the dynamics.

4.3.2. Parameter estimates

We now discuss the results from the selected Cprobit-RW model. As is usual for Bayesian inference, we summarize the posterior distributions of the parameters by reporting their posterior means and posterior standard deviations. Figure 2 shows the time evolution of the state vectors $\boldsymbol{\mu}_t$. It is clear from this figure that there are significant dynamics in preferences. The population mean for the price coefficient decreased in value over time passing approximately from −0.05 to −0.18. This clearly indicates that consumers are becoming increasingly price sensitive over the time span of our data. This suggests that while a regular price of one brand may be acceptable to consumers on a given choice occasion, the same regular price may seem too high on later occasions. From an economic theory perspective, this result suggests that optimal prices should be decreasing over time. In addition to the increased competition between brands, one possible explanation for this increased price sensitivity over time may be consumers' frequent exposure to price promotions which may lower their reference prices (e.g., Kalyanaraman and Winer, 1995). This result is consistent with those obtained by Jedidi et al. (1999) who used a distributed-lag specification to model consumers's sensitivities.

Figure 2 shows that the promotion sensitivity is more or less stable over time.[4] Similarly, the brand-specific intercepts appear to fluctuate in a non-systematic way over the eight-year period. The positive (negative) sign of brand j's intercept ($j = 1, 2, 3$) indicates that on average households have a higher (lower) base utility for brand j than for the reference brand (brand four). The changes of intercepts over time reflect residual fluctuations in brand share differences over time after controlling for the effects of price and promotion variables. As brand intercepts are measures of base choice probabilities, this result suggests that the four brands have been successful at maintaining their market share positions over the long run

[4] We cannot explain the dip in promotion sensitivity around period 50 based on the data we have. Specifically, we do not have any information that suggests that a structural change occurred during the data period, for example, there were no radical changes in scanner technology. In addition, Figure 1 does not indicate any drastic changes in the pricing and promotions policies.

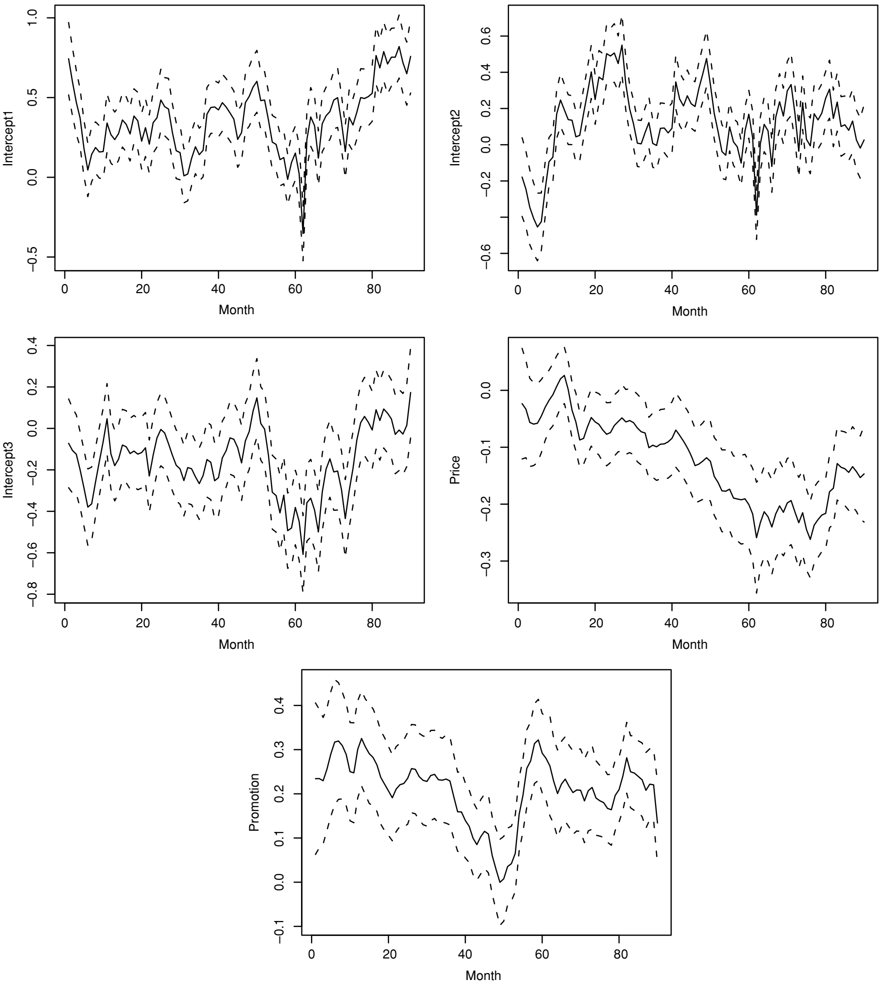

Fig. 2 Evolution of brand intercepts and marketing variable sensitivities over time (91 months) in the Bayesian state-space, heterogeneous multinomial probit model (Cprobit-RW). Posterior means (solid lines) and one-standard deviation posterior intervals (dashed lines).

and any gains in share are only short-lived (DeKimpe and Hanssens, 1995). See also Figure 1. Overall, the results in figure 2 suggest that while market shares and promotion sensitivity are more or less stable in the long run, consumers have become more price sensitive.

Table 2 focuses on characterizing the magnitude of unobserved heterogeneity and dynamics. The top panel reports the posterior means and the corresponding posterior standard deviations for the elements of $\boldsymbol{\Psi}$. Recall that $\boldsymbol{\Psi}$ represents the uncertainty in the time-varying random shocks that influence the transition equation. Most of the diagonal elements of $\boldsymbol{\Psi}$ are small in magnitude but large relative to their corresponding posterior standard deviations, suggesting that the unobserved effects capture important sources of variation in the population means over time. Most of the covariance elements are not "significantly" different from zero. This suggests that the population means vary more or less independently over time.

Table 2 Cross-sectional and time-varying heterogeneity

	Intercept 1	Intercept 2	Intercept 3	Price	Promotion
Ψ					
Intercept 1	0.04	0.03	0.02	0.002	−0.004
	(0.02)	(0.01)	(0.01)	(0.004)	(0.006)
Intercept 2		0.04	0.01	0.002	−0.003
		(0.01)	(0.01)	(0.003)	(0.006)
Intercept 3			0.02	0.0004	−0.004
			(0.01)	(0.003)	(0.005)
Price				0.002	0.0002
				(0.001)	(0.001)
Promotion					0.006
					(0.003)
Ω					
Intercept 1	1.75	0.22	−0.15	0.04	0.03
	(0.24)	(0.15)	(0.29)	(0.03)	(0.05)
Intercept 2		0.29	0.22	0.006	−0.02
		(0.09)	(0.14)	(0.01)	(0.02)
Intercept 3			1.38	−0.009	0.008
			(0.58)	(0.03)	(0.05)
Price				0.005	0.001
				(0.003)	(0.004)
Promotion					0.01
					(0.006)

[1]Posterior mean and standard deviations (in parenthesis) of $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$

The bottom panel of Table 2 shows the posterior means and the corresponding posterior standard deviations for the elements in $\boldsymbol{\Omega}$. Recall that $\boldsymbol{\Omega}$ is the common covariance matrix for the population distributions. It characterizes the magnitude of the unobserved cross-sectional heterogeneity. The pattern of the covariance results are the same as in $\boldsymbol{\Psi}$. Comparing the top and bottom panels of Table 2, it is clear that the variance associated with consumer heterogeneity is greater than that associated with dynamics. In fact, all the diagonal elements of $\boldsymbol{\Omega}$ are larger than those of $\boldsymbol{\Psi}$. This result is consistent with the model selection results in Table 1. Figure 3 shows the histograms of the posterior means for the household-specific parameters $\mathbf{b}_h$. Recall that all of these coefficients are centered at zero and therefore are deviations from the time varying population mean $\boldsymbol{\mu}_t$.

Finally, the correlation coefficients, $\mathbf{R} = \{\{r_{ij}\}\}$, between the utilities of brand i and j $(i, j = 1, 2, 3,\ i \neq j)$ are all positive with posterior means ($r_{12} = 0.436,\ r_{13} = 0.278,\ r_{23} = 0.618$). The positive correlation among the utility errors reflects the fact that the errors are all differenced with respect to the base alternative. The free standard deviations within $\boldsymbol{\Sigma}$ have posterior means $\sigma_2 = 0.474$ and $\sigma_3 = 0.763$.

4.4. Elasticities

Assessing the effects of changes in marketing policies (e.g, price and promotion) on brand shares has significant implications for retailers and manufacturers. For example, differences

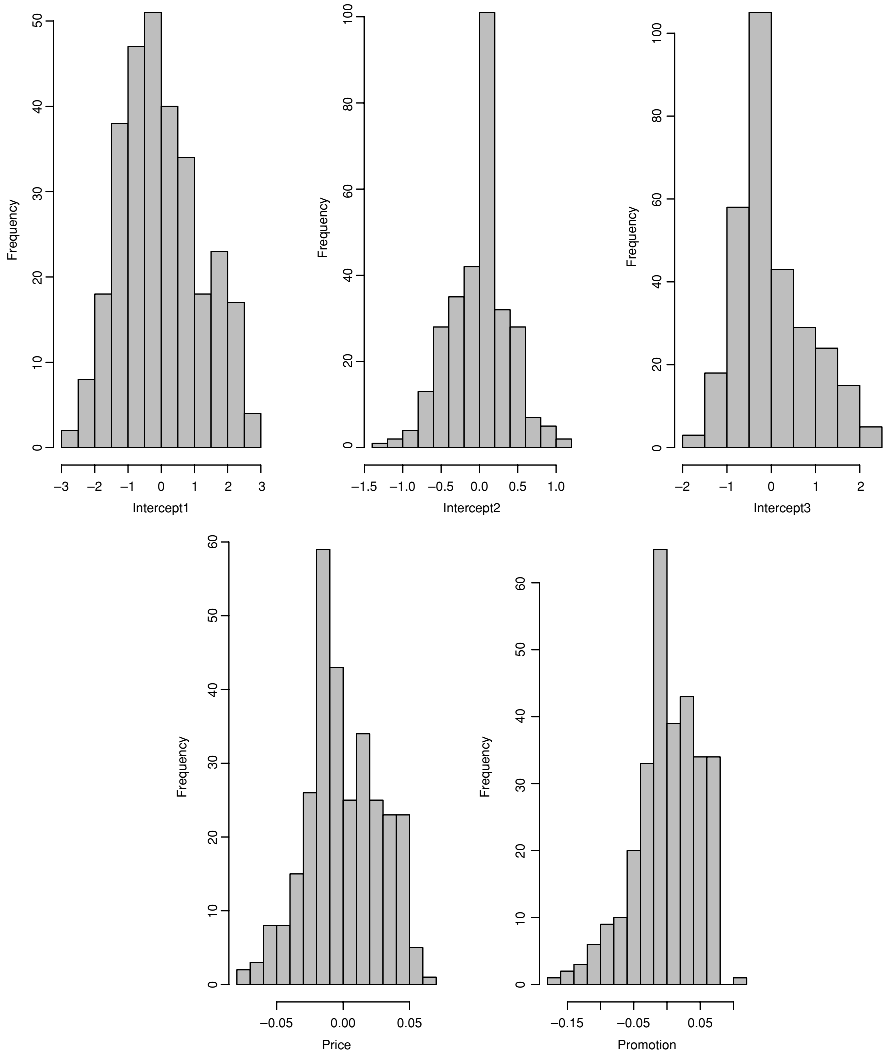

Fig. 3 Histograms of the posterior means of the household-specific parameters $\mathbf{b}_h$ in the Bayesian state-space, heterogeneous multinomial probit (Cprobit-RW) model

in price elasticities over time suggest different dynamic pricing policies across brands. We conducted some market simulations to evaluate these effects. Specifically, we computed the price and promotion elasticities of brand shares. In these simulations, we first calculated the predicted choice probabilities in the absence of any changes in marketing policies. These predicted choice probabilities give the base choice probabilities against which the effects of changes in marketing policies can be evaluated.

We used the MCMC draws of the parameters and our data to compute the choice probabilities for each brand for each household and each purchase occasion. We computed the base choice probabilities for the selected Cprobit-RW and each of the three null models (Sprobit, Hprobit, and Tprobit) to compare the price and promotion elasticities across these models. Contrasting the difference among the elasticities from the different models allows us to understand the effects of ignoring the brand choice dynamics, expressed in terms of time-varying parameters, on the market responses.

To assess the impact of a regular price change by a brand, we increased the price of the target brand by 1% while holding the other variables constant. The probabilities of choosing the brands were recomputed using the manipulated price for all purchases. This produces new choice probabilities for the different brands at each purchase occasion and for each household. The percent change in choice probabilities divided by the percent change in price is taken as the regular price elasticity. The promotion elasticities were computed similarly.

To estimate the aggregate elasticity for a given brand, we took the average of all elasticities for that brand across all observations. To compute the temporal elasticities, we took the average of the elasticities across all observations within each month.

4.4.1. Aggregate price elasticities

Table 3 reports the aggregate regular price elasticities. The entries in the table represent the direct and cross elasticities from the selected Cprobit-RW model and the three null models. All direct price elasticities have the expected negative sign and all cross-price elasticities are positive as these brands are essentially substitutes. Across brands, the average price elasticity from Cprobit-RW is −1.22 which suggests that consumers are quite sensitive to price. The average cross-price elasticity is 0.485 which suggests that brands are vulnerable to switching in this category.

Brand 1's own-price choice elasticity from Cprobit-RW is −0.63. Therefore, a 1% increase of brand 1's price results in a 0.63% decrease in Brand 1's share. This same price increase leads to a 0.60, 0.44, and 0.50% increase in the share of Brand 2, 3, and 4, respectively. Brand 1, the market leader, has the lowest price elasticity and Brand 2 has the highest price elasticity, suggesting that these brands are the least and most vulnerable, respectively, to brand switching.

Table 3 reveals important differences in the magnitudes of the own and cross price elasticities across the four models. Specifically, the simple probit model (Sprobit) which ignores both preference dynamics and consumer unobserved heterogeneity has uniformly lower magnitude elasticities when compared to Cprobit-RW. In fact, the average (across brands) own price elasticities from the Cprobit-RW model (−1.22) is larger in magnitude than that from the Sprobit model (−0.95). The pattern of bias created from ignoring either consumer heterogeneity or preference dynamics is not as systematic. For Brand 3, all the direct and cross-price elasticities from Cprobit-RW are all larger in magnitude than those from Hprobit and Tprobit. This result does not hold true, however, for Brands 1, 2, and 4. As Cprobit-RW has the greatest support from the data (based on the marginal log likelihood), the outcome that the elasticities are different across models, suggests that failure to account for both preference dynamics and consumer heterogeneity is likely to result in misleading inferences about the magnitude of the direct and cross price elasticities.

4.4.2. Temporal price elasticities

Figure 4 contrasts the time-varying price elasticities for the four brands obtained from the four models. Each panel in Figure 4 represents the temporal variation of the own-price elasticities from the four models for a given brand. The temporally varying price elasticities are computed by averaging the price elasticities computed from the observations pertaining to that month.

Table 3 Aggregate direct and cross price elasticities

Elasticity	Sprobit	Hprobit	Tprobit	Cprobit-RW
$1 \to 1$	−0.34	−0.29	−0.56	−0.63
$1 \to 2$	0.35	0.36	0.52	0.60
$1 \to 3$	0.21	0.18	0.25	0.44
$1 \to 4$	0.39	0.22	0.81	0.50
$2 \to 1$	0.11	0.12	0.15	0.23
$2 \to 2$	−1.48	−1.56	−1.14	−1.63
$2 \to 3$	0.35	0.85	0.16	0.59
$2 \to 4$	0.67	0.85	0.58	0.69
$3 \to 1$	0.08	0.08	0.10	0.23
$3 \to 2$	0.46	0.66	0.22	0.69
$3 \to 3$	−0.76	−1.25	−0.56	−1.27
$3 \to 4$	0.18	0.32	0.13	0.44
$4 \to 1$	0.12	0.08	0.27	0.21
$4 \to 2$	0.67	0.84	0.67	0.83
$4 \to 3$	0.14	0.33	0.11	0.37
$4 \to 4$	−1.21	−1.22	−1.42	−1.34

Ex: Row $1 \to 2$ gives the cross-price elasticity of 1's price on 2's share.

For the Cprobit-RW model, we can see that the temporal price elasticities mimic the pattern observed in the price coefficient in Figure 2. For each brand, we see that the temporal price elasticities for Tprobit and Cprobit-RW exhibit considerable decrease over the span of the data. In contrast, for the models that ignore temporal variation in preferences, these elasticities remain fairly constant over time with some short-term fluctuations. This difference in pattern is mostly explained by the variation of the price parameters in the state-space models.

Note that the own price elasticity or cross price elasticity is generally a function of the price parameter, the price level, and the probability of choice. Thus variability in price elasticities can be attributed to changes in the price coefficients and changes in price levels. Since the price coefficient is invariant over time in Sprobit and Hprobit, most of the price elasticity fluctuations are therefore due to changes in the price levels. In contrast, the price elasticity fluctuations in Tprobit and Cprobit are due to changes in both the price levels and the price coefficients. Thus, the static multinomial probit model (Sprobit) and the hierarchical multinomial probit model (Hprobit), which are standard models of choice in the literature have limited flexibility in capturing the evolution of choice behavior. Including cross-sectional heterogeneity, while important, does not appear sufficient in modeling the variation in preferences and choices. The elasticity results suggest that incorporating preference dynamics is important for adequately assessing the market response to marketing actions.

4.4.3. Aggregate promotion elasticities

In Table 4 we present aggregate, own and cross promotion elasticities for the four models. All direct promotion elasticities have the expected positive sign, whereas the cross promotion

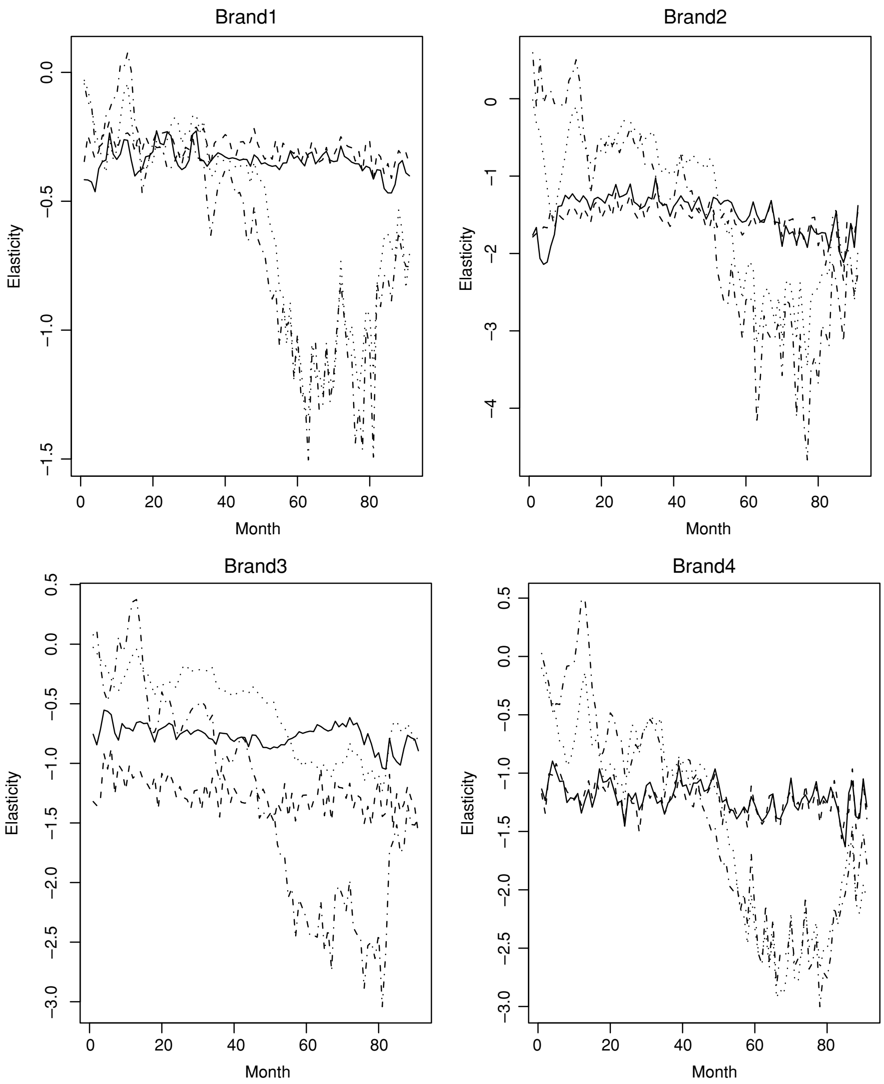

Fig. 4 Temporal direct (own) price elasticities of brand shares for Sprobit (solid line), Hprobit (dashed line), Tprobit (dotted line) and Cprobit-RW (dashed-dotted line) models

elasticities are all negative. The signs of the own and cross promotion elasticities are consistent with the fact that the different brands within the category are all substitutes. A comparison of the magnitudes of the direct and cross promotion elasticities across the models reveals that all four models generate essentially the same promotion elasticities. Most notable are the differences between the promotion elasticities from the Cprobit-RW model and those from the Hprobit model. This similarity across models can be understood within the context of the pattern observed in Figure 2. There, we saw that the promotion parameter did not exhibit any appreciable dynamics in the population. The lack of dynamics in the population mean for the promotion parameter essentially implies that the four models are similar in their ability to handle promotion effects.

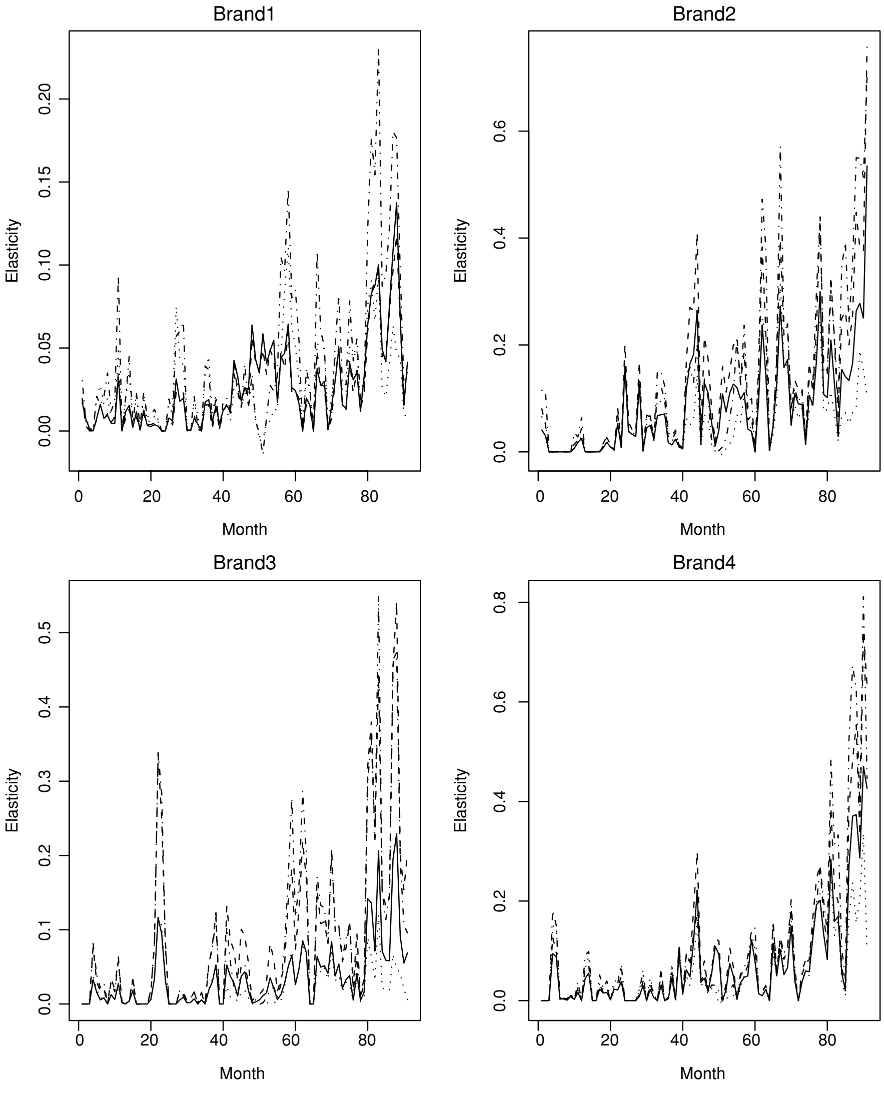

Fig. 5 Temporal direct (own) promotion elasticities of brand shares for Sprobit (solid line), Hprobit(dashed line), Tprobit (dotted line) and Cprobit-RW (dashed-dotted line) models

Brand 2's direct promotion elasticity from the Cprobit-RW model is 0.12 (see Table 4). Therefore, a 1% increase in Brand 2's promotion results in a 0.12% increase in Brand 2's share. The same promotion increase leads to a 0.02%, 0.04%, and 0.05% decrease in the share of Brands 1, 3, and 4, respectively. Among all brands, Brand 1 (Brand 2) appears to be the least (most) sensitive to promotion changes, but differences across brands are minimal. Across brands, the average promotion elasticity from the Cprobit-RW model is 0.0825 and the average cross promotion elasticity is 0.04. Comparing the promotion and price elasticities from the general model, we find that, on average, the magnitude of aggregate promotion elasticities is much smaller relative to that of regular price elasticities. This significantly

Table 4 Aggregate direct and cross price elasticities

Elasticity	Sprobit	Hprobit	Tprobit	Cprobit-RW
$1 \to 1$	0.03	0.02	0.03	0.04
$1 \to 2$	−0.03	−0.04	−0.04	−0.06
$1 \to 3$	−0.02	−0.03	−0.02	−0.06
$1 \to 4$	−0.03	−0.03	−0.06	−0.06
$2 \to 1$	−0.01	−0.01	−0.01	−0.02
$2 \to 2$	0.08	0.12	0.06	0.12
$2 \to 3$	−0.03	−0.08	−0.01	−0.04
$2 \to 4$	−0.06	−0.08	−0.04	−0.05
$3 \to 1$	−0.01	−0.01	−0.01	−0.01
$3 \to 2$	−0.03	−0.04	−0.01	−0.03
$3 \to 3$	0.03	0.08	0.02	0.08
$3 \to 4$	−0.01	−0.02	−0.01	−0.03
$4 \to 1$	−0.01	−0.01	−0.02	−0.02
$4 \to 2$	−0.06	−0.07	−0.04	−0.06
$4 \to 3$	−0.01	−0.04	−0.01	−0.03
$4 \to 4$	0.06	0.08	0.06	0.09

Ex: Row $1 \to 2$ gives the cross elasticity of 1's promotion on 2's share.

higher response to price compared to promotion suggests that, in this product category, pricing policy has a higher impact on consumers' buying process than short-term promotion policy.

4.4.4. Temporal promotion elasticities

Figure 5 displays the estimated temporal own promotion elasticities obtained from the four different models. From the figure, we can see how the promotion elasticities implied by the four models are spread out over time. The own promotion elasticities for all brands are positive with an upward trend. The temporal pattern of promotion elasticities is similar across models. This similarity is consistent with the results for the aggregate elasticities in Table 4. From Figure 5, we observe that the promotion elasticities increased in the later portion of the data. This variation in elasticities mimics the variation in the promotion spending displayed in Figure 1 where we see that all four brands increased their promotion spending in the last 10 months. The increase in elasticity reflects this increased level of spending in the data, and not any significant differences in the promotion parameter over time.

5. Conclusion

Understanding choice and preference dynamics is a major issue in the choice modeling literature. Variation in preferences and choices can occur due to changes in marketing variables, changes in response coefficients, and changes in serially correlated error terms. Previous models focused mostly on the changes in the values of independent variables and on the serial correlation of the error terms. In this paper, we modeled preference variation through explicit modeling of changes over time in the model parameters. We considered various forms of parameter dynamics that are special cases of our general model. These include a random walk process, dynamic random effects, and a diagonal transition matrix. We developed a hierarchical Bayesian state-space model to allow for dynamically-linked population distributions and an estimation approach that simultaneously allows for time-varying parameters and cross-sectional heterogeneity across households. We showed that the proposed model is very general and nests traditional models such as the static multinomial probit and the hierarchical multinomial probit model.

We estimated seven competing models using over eight years of scanner panel data for 300 households for a consumer packaged good. Model comparison results pointed to a heterogenous, state-space model with a random walk process. Thus, in addition to being heterogeneous across consumers, the model parameters do indeed vary over time. In particular, we found that consumers became more and more price sensitive over the time span of the data as revealed by the increasing magnitude of price elasticities over time. This result may be due to consumers's frequent exposure to price promotions which can lower their reference price. Our results also show that null models (static and heterogeneous) that do not fully capture the dynamics of choice behavior (variability due to changes in coefficients) can distort the magnitude and the pattern of price elasticities over time. The positive performance of our model and the fact that most marketing datasets involve observations over many time periods lead us to believe that the state space approach can be gainfully employed in routinely analyzing longitudinal data in marketing.

Future research can extend our model in many different ways. In this paper, we assumed a state-space, VARX(1) process for the parameter dynamics. Our model cannot therefore capture non-linear changes (e.g., step functions, parabolic functions) that might occur in certain contexts. A possible extension then would be to model parameter dynamics with different stochastic structures (dynamic simultaneous equations, VAR(p) models with $p > 1$). The last model is of particular interest as it allows one to model directly the dependence among the coefficients across multiple time periods. While we focused on VARX(1) process, it is important to note that the state space approach is very flexible as it can incorporate other more general forms of dynamics. Instead of a state space approach, a hidden markov or change-point framework could be adopted as an alternative to accommodate discrete changes and structural shifts in the parameters. A comparison of these alternative would be interesting to gauge their ability to model preference evolution.

In modeling the dynamics, we allowed the population mean to vary over time, thus generating a dynamic sequence of population distributions. This approach can be extended to allow the population variance to be dynamic, again using a state-space framework. In our model, as only the population mean is allowed to evolve, the coefficients of every household are forced to follow the same pattern over time, which may not be realistic in some situations. Finally, various marketing phenomena are missing in our research. The present model can be extended to incorporate incidence, quantity and timing decisions (Chintagunta,1993; Pauwels et al., 2002) of the households. In addition, our approach can be used in models that capture the endogeneity of marketing variables (Yang et al., 2003). Here, dynamics can be incorporated both in the demand and supply side specifications. We leave these extensions to future research.

Appendix: Priors and full conditional distributions

In this Appendix, we describe the priors and full conditional distributions for the parameters associated with the model described in Eqs. (11) to (16). We begin with a description of the priors.

The priors for the utilities, time-varying parameters and the individual-specific random effects are directly available from the model specification. For the other parameters, we specify independent and diffuse (but proper) priors. The prior for the precision matrix, $\mathbf{\Psi}^{-1}$, is Wishart, $W[\rho_1, (\rho_1\mathbf{R}_1)^{-1}]$, where $\rho_1 \geq L$ is the "degree of freedom" parameter for the Wishart and $\mathbf{R}_1$ is a $L \times L$ scale matrix. Given our parameterization of the Wishart, $\mathbf{R}_1$ is the expected prior conditional variance of the time-varying parameters, $\boldsymbol{\mu}_t$'s. Similarly, the prior for the precision matrix $\mathbf{\Omega}^{-1}$ is $W[\rho_2, (\rho_2\mathbf{R}_2)^{-1}]$, where $\rho_2 \geq L$ and $\mathbf{R}_2$ is a $L \times L$ scale matrix. We set $\rho_1 = \rho_2 = L + 1$ and $\mathbf{R}_1 = \mathbf{R}_2 = \text{diag}(0.01, 0.01, 0.01, 0.002, 0.004)$. The dynamics are initialized by assuming that the prior for initial population mean $\boldsymbol{\mu}_0$ is multivariate normal $N(\boldsymbol{\mu}_0, \mathbf{C}_0)$. We set $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{C}_0 = 100\mathbf{I}$ to obtain a diffuse prior. The prior for $\boldsymbol{\theta}$ is assumed to be multivariate normal $\text{N}(\mathbf{d}_\theta, \mathbf{D}_\theta)$, where $\mathbf{d}_\theta = \mathbf{0}$ and $\mathbf{D}_\theta = 100\mathbf{I}$. Let $\boldsymbol{\delta} = \text{vec}(\mathbf{\Phi})$ be the vector of all the elements in $\mathbf{\Phi}$. We assume a normal prior $N(\mathbf{d}_\delta, \mathbf{D}_\delta)$, for $\boldsymbol{\delta}$, where $\mathbf{D}_\delta$ is a diagonal matrix with its diagonal elements set to 5. The elements in $\mathbf{d}_\delta$ corresponding to the off-diagonal elements in $\mathbf{\Phi}$ are set to zero and those corresponding to the diagonal elements in $\mathbf{\Phi}$ are set at one.

The variance-covariance matrix $\mathbf{\Sigma}$ is constrained as one of its variance elements is fixed at one. Many different approaches have been proposed in the literature for specifying priors for $\mathbf{\Sigma}$. For example, McCulloch and Rossi (1994) specified priors on the unidentified covariance matrix. More recently, McCulloch et al. (2000) and Nobile (2000) describe how to directly specify priors on the identified elements of the covariance matrix. Here, we follow Barnard et al. (2000) in setting the prior in terms of the standard deviations and correlations in $\mathbf{\Sigma}$. The covariance matrix $\mathbf{\Sigma}$ can be decomposed into a correlation matrix, $\mathbf{R}$ and a vector $\mathbf{s}$, of standard deviations, i.e., $\mathbf{\Sigma} = \text{diag}(\mathbf{s}) \times \mathbf{R} \times \text{diag}(\mathbf{s})$, where $\mathbf{s} = (\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{JJ}})'$. Let $\boldsymbol{\omega}$ contain the logarithms of the elements in $\mathbf{s}$. We assume a multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$ for the non-redundant elements of $\mathbf{R}$, such that it is constrained to the subspace of the $M * (M-1)/2$ dimensional cube $[-1, 1]^{M(M-1)/2}$ that yields a positive definite correlation matrix. Finally, we assume that the log-standard deviations in $\boldsymbol{\omega}$ come from independent standard normal priors.

We now focus on the full conditional distributions for the unknowns in the model.

1. The full conditional distribution for the household-specific random effects, $\mathbf{b}_h$, $h = 1, \ldots, H$ can be obtained from standard Bayesian arguments involving conjugate distributions. The full conditional is multivariate normal, $N(\mathbf{b}_h^*, \mathbf{V}_{bh})$, where $\mathbf{b}_h^* = \mathbf{V}_{bh}(\sum_{j=1}^{n_h} \mathbf{X}_{hj}' \mathbf{\Sigma}^{-1} \tilde{\mathbf{u}}_{hj})$ and $\mathbf{V}_{bh}^{-1} = \sum_{j=1}^{n_h} \mathbf{X}_{hj}' \mathbf{\Sigma}^{-1} \mathbf{X}_{hj} + \mathbf{\Omega}^{-1}$ and $\tilde{\mathbf{u}}_{hj} = \mathbf{u}_{hj} - \mathbf{X}_{hj}\boldsymbol{\mu}_{t_{hj}}$ is a vector of adjusted utilities.
2. To obtain the vector of states $\{\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T\}$, we use the forward-filtering, backward-sampling algorithm. Alternative methods include one at a time sampling procedures, such as those outlined in Tsay and McCulloch (1994). The forward-filtering, backward sampling algorithm consists of two steps. In a forward step, the moments of the updated distribution of each state is computed using a Kalman filter approach. In the backward step, each parameter is sampled from its conditional distribution conditioned on the preceding draw. We begin by defining the model in terms of the time periods $t = 1$ to T.

$$\begin{aligned} \widetilde{\mathbf{u}}_t &= \mathbf{X}_t\boldsymbol{\mu}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t), \quad \mathbf{\Sigma}_t = \mathbf{\Sigma} \otimes \mathbf{I}_{n_t} \\ \boldsymbol{\mu}_t &= \boldsymbol{\theta} + \mathbf{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{\Psi}), \end{aligned}$$

where n_t is the number of observations in period t, $\widetilde{\mathbf{u}}_t$ is the $n_t \times M$ vector of adjusted utilities stacked observation by observation and obtained after subtracting the systematic component involving the random effects from the utilities and $\mathbf{X}_t$ is obtained by vertically stacking the $\mathbf{X}_{hj}$ matrices for all observations belonging to time-period t.

In the forward step, the moments of each state are computed recursively as follows: Let the posterior at time $t-1$ be $p(\boldsymbol{\mu}_{t-1}|D_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$, where D_{t-1} is the information set at time-period $t-1$ and includes the other unknowns and the utilities upto time period $t-1$. The prior for $\boldsymbol{\mu}_t$ can be written as $p(\boldsymbol{\mu}_t|D_{t-1}) \sim N(\gamma_t, \Gamma_t)$, where, $\gamma_t = \boldsymbol{\theta} + \boldsymbol{\Phi}\mathbf{m}_{t-1}$ for $t > 1$, $\gamma_1 = \boldsymbol{\theta} + \boldsymbol{\Phi}\boldsymbol{\eta}_0$, and $\Gamma_t = \boldsymbol{\Phi}\mathbf{C}_{t-1}\boldsymbol{\Phi}' + \boldsymbol{\Psi}$. After the data for period t are observed, conditional on the utilities for the observations in period t, the posterior at time t can be written as $p(\boldsymbol{\mu}_t|D_{t-1}, \widetilde{\mathbf{u}}_t) \sim N(\mathbf{m}_t, \mathbf{C}_t)$ where $\mathbf{C}_t^{-1} = \Gamma_t^{-1} + \mathbf{X}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{X}_t$ and $\boldsymbol{\mu}_t = \mathbf{C}_t[\Gamma_t^{-1}\gamma_t + \mathbf{X}_t'\boldsymbol{\Sigma}_t\widetilde{\mathbf{u}}_t]$. These posterior moments can be computed and stored in the forward step of the algorithm.

Let $\widetilde{\boldsymbol{\mu}} = \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T\}$ and let the parameters that are not in $\widetilde{\boldsymbol{\mu}}$ be written as $\widetilde{\boldsymbol{\alpha}}$. Then we can write $p(\widetilde{\boldsymbol{\mu}}|D_T, \widetilde{\boldsymbol{\alpha}}) = p(\boldsymbol{\mu}_T \mid D_T, \widetilde{\boldsymbol{\alpha}})p(\boldsymbol{\mu}_{T-1} \mid \boldsymbol{\mu}_T, D_{T-1}, \widetilde{\boldsymbol{\alpha}}) \ldots p(\boldsymbol{\mu}_0 \mid \boldsymbol{\mu}_1, D_0, \widetilde{\boldsymbol{\alpha}})$. This depends upon the identity $p(\boldsymbol{\mu}_{T-k} \mid \boldsymbol{\mu}_{T-k+1}, D_T, \widetilde{\boldsymbol{\alpha}}) = p(\boldsymbol{\mu}_{T-k}|\boldsymbol{\mu}_{T-k+1}, D_{T-k}, \widetilde{\boldsymbol{\alpha}})$. Therefore, $\widetilde{\boldsymbol{\mu}}$ can be sampled using the following steps

1. Sample $\boldsymbol{\mu}_T$ from $(\boldsymbol{\mu}_T|D_T, \widetilde{\boldsymbol{\alpha}}) \sim N(\mathbf{m}_T, \mathbf{C}_T)$
2. for $t = T-1, \ldots, 0$, sample $\boldsymbol{\mu}_t$ from $p(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t+1}, D_t, \widetilde{\boldsymbol{\alpha}}) = N(\mathbf{q}_t, \mathbf{Q}_t)$.

In the second step, we use $\mathbf{Q}_t^{-1} = \mathbf{C}_t^{-1} + \boldsymbol{\Phi}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}$ and $\mathbf{q}_t = \mathbf{Q}_t[\mathbf{C}_t^{-1}\mathbf{m}_t + \boldsymbol{\Phi}'\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu}_{t+1} - \boldsymbol{\theta})]$, and for the initial state we have $p(\boldsymbol{\mu}_0|\boldsymbol{\mu}_1, D_0, \widetilde{\boldsymbol{\alpha}}) = N(\mathbf{q}_0, \mathbf{Q}_0)$ where, $\mathbf{Q}_0^{-1} = \mathbf{C}_0^{-1} + \boldsymbol{\Phi}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}$ and $\mathbf{q}_0 = \mathbf{Q}_0[\mathbf{C}_0^{-1}\boldsymbol{\eta}_0 + \boldsymbol{\Phi}'\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\theta})]$.

3. The full conditional distribution for $\boldsymbol{\Psi}^{-1}$ is Wishart and is obtained by combining the normal likelihood for the time-varying parameters with the Wishart prior distribution for $\boldsymbol{\Psi}^{-1}$. Given the conjugacy of the Wishart with the normal likelihood, the posterior is given by $W(\rho_1 + T, [\sum_{t=1}^{T}(\boldsymbol{\mu}_t - \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} - \boldsymbol{\theta})(\boldsymbol{\mu}_t - \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} - \boldsymbol{\theta})' + \rho_1\mathbf{R}_1]^{-1})$, where T is the number of periods.

4. The full conditional for $\boldsymbol{\theta}$ is a normal $N(\hat{\boldsymbol{\theta}}, \mathbf{V}_\theta)$. Let $\tilde{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_t - \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1}$, then $\mathbf{V}_\theta = \mathbf{D}_\theta^{-1} + T\boldsymbol{\Psi}^{-1}$, and $\hat{\boldsymbol{\theta}} = \mathbf{V}_\theta[\mathbf{D}_\theta^{-1}\mathbf{d}_\theta + \sum_{t=1}^{T}\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{\mu}}_t]$.

5. The full conditional for $\boldsymbol{\delta}$ is a normal $N(\hat{\boldsymbol{\delta}}, \mathbf{V}_\delta)$. Let $\tilde{\boldsymbol{\mu}}_{t\delta} = \boldsymbol{\mu}_t - \boldsymbol{\theta}$, and let $\mathbf{W}_t$ be a $L \times (L \times L)$ matrix generated from using $\boldsymbol{\mu}_{t-1}$ on each row, such that $\tilde{\boldsymbol{\mu}}_{t\delta} = \mathbf{W}_t\boldsymbol{\delta} + \boldsymbol{\epsilon}_t$ forms a SUR system. Then, $\mathbf{V}_\delta = \mathbf{D}_\delta^{-1} + \sum_{t=1}^{T}\mathbf{W}_t'\boldsymbol{\Psi}^{-1}\mathbf{W}_t$, and $\hat{\boldsymbol{\delta}} = \mathbf{V}_\delta[\mathbf{D}_\delta^{-1}\mathbf{d}_\delta + \sum_{t=1}^{T}\mathbf{W}_t'\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{\mu}}_{t,\delta}]$.

4. The full conditional distribution for $\boldsymbol{\Omega}^{-1}$ is obtained in the same manner as for $\boldsymbol{\Psi}^{-1}$ by combining the normal likelihood for the household-specific parameters with the prior distribution for $\boldsymbol{\Omega}^{-1}$. It is given by $W\left(\rho_2 + H, [\sum_{h=1}^{H}\mathbf{b}_h\mathbf{b}_h' + \rho_2\mathbf{R}_2]^{-1}\right)$ where H is the number of households.

5. The full conditional distribution for each free element in the vector of log standard deviations $\boldsymbol{\omega}$ of the utility errors can only be written up-to a normalizing constant (recall that the first term in $\boldsymbol{\omega}$ is fixed to 0 (i.e., $\sigma_1 = 1$) for identification purpose). Given our assumption of an independent normal prior for each free element, we use a Metropolis - Hastings step to simulate independently each free element in $\boldsymbol{\omega}$. A univariate normal proposal density can be used to generate candidates for this procedure. If $\omega_i^{(m-1)}$ is the current value of ith component of $\boldsymbol{\omega}$, then a candidate value is generated using a random walk chain $\omega_i^c = \omega_i^{(m-1)} + N(0, \tau)$, where τ is a tuning constant that controls the acceptance rate. The proposed candidate ω_i^c is accepted as the new value $\omega_i^{(m)}$ with probability,

$$\min\left\{\frac{L\left(\omega_i^c\middle|\{\{\mathbf{u}_{hj}\}\}, \{\boldsymbol{\mu}_t\}, \{\mathbf{b}_h\}, \mathbf{R}, \boldsymbol{\omega}_{-i}\right)p\left(\omega_i^c\right)}{L\left(\omega_i^{(m-1)}\middle|\{\{\mathbf{u}_{hj}\}\}, \{\boldsymbol{\mu}_t\}, \{\mathbf{b}_h\}, \mathbf{R}, \boldsymbol{\omega}_{-i}\right)p\left(\omega_i^{(m-1)}\right)}, 1\right\}. \tag{17}$$

In the above acceptance probability expression, $L(\omega_i^c \mid \{\{\mathbf{u}_{hj}\}\}, \{\boldsymbol{\mu}_t\}, \{\mathbf{b}_h\}, \mathbf{R}, \boldsymbol{\omega}_{-i})$ is the likelihood of observing the latent utilities, evaluated at the candidate value ω_i^c and the quantity $p(\omega_i^c)$ represents the prior univariate normal density evaluated at ω_i^c. If the candidate is rejected, then $\omega_i^{(m)} = \omega_i^{(m-1)}$.

6. Many different approaches can be used to sample the correlation matrix **R**. Here, we use slice sampling (Neal, 2003) to sample each non-redundant correlation in **R**, separately. The full conditional for an arbitrary correlation r is proportional to the function $f(r) = L(r)p(r)$, where $L(r)$ is the likelihood of the utilities evaluated at the value r and $p(r)$ is the prior density. In slice sampling, the current value r^c (i.e, the value from the previous iteration of the MCMC sampler) is replaced with a value r^m using the following steps

- Draw a value z uniformly from $(0, f(r^c))$ and define a region (called the Slice) $M = \{r : z < f(r)\}$ containing r^c.
- Generate an interval around r^c that ensures that **R** is positive definite. We use the computations outlined in Barnard et al. (2000) to derive this interval.
- Draw a new value r^m from the part of the slice m within the interval. In this step, we use the shrinkage procedure outlined in Neal (2003) to successively shrink the interval. The shrinkage procedure results in an efficient approach for generating the new value.

7. The full conditional distribution associated with the identified set of utilities $\mathbf{u}_{hj}$ in the proposed model is a M-variate normal distribution $N(\mathbf{X}_{hj}\boldsymbol{\beta}_{hj}, \boldsymbol{\Sigma})$ truncated over a multidimensional cone (see McCulloch and Rossi (1994) for the case of the traditional multinomial probit model). If the chosen brand is brand m, $y_{hjm} = 1$, then $u_{hjm} > \max(\mathbf{u}_{h,j,-m}, 0)$; If $y_{hjm} = 0$, then $u_{hjm} < \max(\mathbf{u}_{h,j,-m}, 0)$, where $\mathbf{u}_{h,j,-m}$ is a $(M-1)$ dimensional vector of all the components of $\mathbf{u}_{hj}$ excluding u_{hjm}.
Sampling these identified set of utilities $\mathbf{u}_{hj}$ can be done observation by observation in a data augmentation step as in Albert and Chib (1993) and McCulloch and Rossi (1994) by sampling each component of $\mathbf{u}_{hj}$ in a mini-Gibbs sampling step. The truncated multivariate normal draw for $\mathbf{u}_{hj}$ can be pieced together from the truncated conditional univariate normal draws of each component of $\mathbf{u}_{hj}$.

Acknowledgments The authors thank Professor Carl Mela, Duke University, for providing the data used in the empirical application.

References

Albert, J., & Chib, S. (1993). Bayesian analysis of binary and polychotomous data. *Journal of the American Statistical Association, 88,* 667–679.

Allenby, G. M., & Lenk, P. J. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association, 89*(428), 1218–1231.

Allenby, G. M., & Lenk, P. J. (1995). Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice. *Journal of Business and Economic Statistics, 13*(3), 281–289.

Allenby, G. M., & Rossi, P, E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics, 89,* 57–78.

Barnard, J., McCulloch, R., & Meng, X. L. (2000). A natural strategy for modeling covariance matrices with applications to shrinkage. *Statistica Sinica, 10,* 1281–1311

Carter, C. K., & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika, 81,* 541–553.

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

Dekimpe, M., G., & Hanssens, M., D. (1995). The persistence of marketing effects on sales. *Marketing Science, 14,* 1–21.

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

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 398–409.
Geweke, J. F., Keane, M. P., & Runkle, D. E. (1997). Statistical inference in the multinomial probit model. *Journal of Econometrics, 80,* 125–165.
Guadagni, P. M., & Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2*(3), 203–238.
Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika, 57,* 97–109.
Hausman, J., & Wise, D. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica, 45,* 319–339.
Heckman, J. J., Manski, C. F., & McFadden, D. (1981). Statistical models for discrete panel data. *Structural analysis of discrete data with applications*. pp. 114–178, Cambridge: MIT Press.
Jedidi, K., Mela, F. C. & Gupta, S. (1999). Managing advertising and promotion for long-term profitability. *Marketing science, 18,* 1–22.
Kalman, R. E. (1960). A new approach to linear filtering and prediction. *Journal of Basic Engineering, 83,* 95–108.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.
Keane, M. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics, 15*(3), 310–327.
McCulloch, R., & Rossi, P, E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics, 64,* 207–240.
McCulloch, R., Polson, N. G. & Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics, 99,* 173–193.
McCulloch R., & Rossi, P. E. (2000). Reply to nobile. *Journal of Econometrics, 99,* 347–348.
McFadden, D. L. (1973). Conditional logit analysis of qualitative choice behavior, In P. Zarembka (ed.), *Frontiers in Econometrics*, (pp. 105–142), New York: Academic Press.
Mela, F. C., Gupta, S., & Lehmann, L, R. (1997). The long-term impact of promotions and advertising on consumer brand choice. *Journal of Marketing Research, 34,* 248–261.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chem.Phys, 21,* 1087–1091.
Neal, R. M. (2003). Slice sampling (with discussion). *Annals of statistics* pp. 705–767.
Newton, M. A., & Raftery, A. E. (1994). Approximate bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B, 56,* 1–48.
Nobile, A. (2000). Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics, 99,* 335–345.
Pauwels, K., Hanssens, M. D., & Siddarth, S. (2002). The long-term effects of price promotions on category incidence, brand choice and purchase quantity. *Journal of marketing research*.
Priestly, M. B. (1980). State-dependent models: A general approach to non-linear time series analysis. *Journal of Time Series Analysis, 1,* 47–71.
Roy, R., Chintagunta, P. K., & Haldar, S. (1996). A framework for investigating habits, the hand of the past. and heterogeneity in dynamic brand choice. *Marketing Science, 15*(3), 280–299.
Seetharaman, P. B. (2003). Modeling multiple sources of state dependence in random utility models of brand choice: A distributed lag approach. *Marketing Science*, forthcoming.
Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis, 15,* 183–202.
Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, (with discussion). *Journal of the American Statistical Association, 82,* 528–550.
Tsay, R., & McCulloch, R. (1994). Statistical analysis of economic time series via markov switching models. *Journal of Times Series Analysis, 15,* 523–539.
West, M., & Harrison, P. J., (1997). *Bayesian forecasting and dynamic models*, 2nd Edition. New York: Springer-Verlag.
Yang, S., Chen, Y., & Allenby, G. (2003). Bayesian analysis of simultaneous demand and supply. *Quantitative marketing and economics, 1*(3), pp. 251–275.