Editor's Note: This article builds on material from the doctoral dissertation awarded the 2003 Zellner Prize for Best Thesis in Business and Economic Statistics.

Modeling Parametric Evolution in a Random Utility Framework

Jin Gyo KIM
MIT Sloan School of Management, Cambridge, MA 02142 (*jgkim@mit.edu*)

Ulrich MENZEFRICKE
Joseph L. Rotman School of Management, University of Toronto, Toronto, Ontario, Canada, M5S 3E6 (*menzefricke@rotman.utoronto.ca*)

Fred M. FEINBERG
University of Michigan Business School Ann Arbor, MI 48109 (*feinf@umich.edu*)

Random utility models have become standard econometric tools, allowing parameter inference for individual-level categorical choice data. Such models typically presume that changes in observed choices over time can be attributed to changes in either covariates or unobservables. We study how choice dynamics can be captured more faithfully by also directly modeling temporal changes in parameters, using a vector autoregressive process and Bayesian estimation. This approach offers a number of advantages for theorists and practitioners, including improved forecasts, prediction of long-run parameter levels, and correction for potential aggregation biases. We illustrate the method using choices for a common supermarket good, where we find strong support for parameter dynamics.

KEY WORDS: Bayesian model; Choice model; Dynamic model; Logit model; Scanner panel data; Varying-parameter model; Vector autoregressive process.

1. INTRODUCTION

The modeling of sequential, individual-level choice has emerged as a research area of great breadth, with applications throughout economics, statistics, psychology, and elsewhere. A fundamental goal is determining how choices evolve over time, and which variables drive them. A rich literature has emerged to aid researchers in linking exogenous covariates to temporal changes in choices. Because data are available on observed choices but not on unobserved measures of relative "attractiveness" of available options, the dominant method of achieving such a linkage has been through random utility models (McFadden 1973; Manski 1977).

Within the random utility framework, one must specify both a utility and an error structure, as well as some link function to convert to observables. Consequently, the lion's share of research has been dedicated to those tasks. For example, prior approaches to modeling choice dynamics capture temporal changes in individual-level utilities by introducing lagged terms for previous choices (cf. Heckman 1981) or by invoking a generalized stochastic error structure (e.g., Allenby and Lenk 1994, 1995). Although these models do capture some types of changes in utility over time in a systematic way, they do not consider changes in variable weights, and so amount to modeling shifts in the intercept of the deterministic component of utility.

We aim to demonstrate that an essential element of choice or utility dynamics can be captured by directly modeling changes in parameters. To that end, we propose a Bayesian dynamic logit model designed to capture choice dynamics by estimating a vector autoregressive process for the parameters of individuals' linear utility functions. Such an approach allows rigorous investigation of a number of issues of interest in forecasting. First, can parameters be distinguished by whether they are time-varying? If they do evolve, can they be further distinguished by the nature of their evolution? And, most important for prediction, to what extent can understanding the nature of parametric evolution be used to gain superior understanding of future choices? The model that we develop does indeed distinguish parameters along these lines, and uses that knowledge for improved forecasting.

The article is organized as follows. We review earlier literature concerning parameter dynamics, specify a Bayesian model to account for them, and develop methods for its estimation. We then estimate the model on individual-level sequential choices, and demonstrate its in-sample and forecast performance. Finally, we suggest possible sources for such dynamics, as well as potential extensions to the general method.

2. DYNAMIC MODEL SPECIFICATION

2.1 Previous Approaches to Changes in Utilities Over Time

Although we are concerned with statistical issues, we note that numerous behavioral studies have suggested that parameters—as embodied by individual-level sensitivities—do indeed change over time. Research on preference reversals, for example, has demonstrated that the so-called "weight function" of

Journal of Business & Economic Statistics
July 2005, Vol. 23, No. 3
DOI 10.1198/073500104000000550

attributes depends on such contextual features as scale compatibility (Slovic, Griffin, and Tversky 1990), strategy compatibility (Tversky, Sattath, and Slovic 1988), and framing (Kahneman and Tversky 1979; Thaler 1985). Decision weights are also known to be sensitive to the scale change of attribute values in an experimental setting (von Nitzsch and Weber 1993).

Previous statistical approaches to capturing utility change over time can be separated into two broad classes, depending on whether intercept shifts are taken to be deterministic or stochastic. Most previous models that introduce lagged choice variables make use of a deterministic intercept shift; the lagged choice shifts the constant term of the deterministic component of the utility for any particular option, so that the intercept is $\alpha' = \alpha + \gamma D_{t-1}$, where D_{t-1} is a dummy variable for the particular option, equaling 1 only if that option was chosen at time $t-1$. Typically, the effects of the lagged choice variables γ are assumed to be homogeneous across units and options. It is important to note that this approach can only capture intercept shifts (in utilities over time) in a deterministic way. In contrast, Allenby and Lenk (1994, 1995) developed a logistic regression model that updates utilities over time by introducing an autoregressive error structure. In their model, the new intercept α' is given by $\alpha + \rho\varepsilon_{t-1}$, where $0 < |\rho| < 1$ and ε_{t-1} is the stochastic component of utility at time $t-1$. Clearly, neither of these approaches can account for utility changes generated by a change in variable weights over time.

2.2 Observation and Evolution Densities

We describe the dynamic logit model and use three generic subscripts; h denotes an individual unit of observation ($h = 1, \ldots, H$), j denotes an option ($j = 1, \ldots, J$), and t denotes time ($t = 1, \ldots, T$). Let $y_{ht} = j$ denote the event that unit h chooses option j at time t, let $\mathbf{x}_{hjt}$ denote unit h's k-dimensional covariate vector for option j at time t, and let u_{hjt} denote unit h's utility for option j at time t. Thus

$$u_{hjt} = \boldsymbol{\beta}'_{ht}\mathbf{x}_{hjt} + \varepsilon_{hjt}, \tag{1}$$

where $\boldsymbol{\beta}_{ht}$ is a k-dimensional coefficient vector for unit h at time t and ε_{hjt} is an associated error. If ε_{hjt} is iid Gumbel, then a dynamic logit model arises for choice probability p_{hjt} (cf. McFadden 1973),

$$p_{hjt} = p(y_{ht} = j|\boldsymbol{\beta}_{ht}) = \frac{\exp(\boldsymbol{\beta}'_{ht}\mathbf{x}_{hjt})}{\sum_{i=1}^{J}\exp(\boldsymbol{\beta}'_{ht}\mathbf{x}_{hit})}. \tag{2}$$

To model parametric temporal variation, we assume that $\boldsymbol{\beta}_{ht}$ can be decomposed into two parts,

$$\boldsymbol{\beta}_{ht} = \boldsymbol{\beta}_t + \mathbf{b}_h, \tag{3}$$

where $\boldsymbol{\beta}_t$ is a time-varying coefficient vector common across units and $\mathbf{b}_h$ is a vector of random effects to incorporate heterogeneity *across* units. Now (2) simplifies to

$$p_{hjt} = p(y_{ht} = j|\boldsymbol{\beta}_t, \mathbf{b}_h). \tag{4}$$

To capture dynamics for $\boldsymbol{\beta}_t$, we introduce a vector autoregressive process of order p, VAR(p) (see, Li and Tsay 1998; Lütkepohl 1991; Polasek and Kozumi 1996),

$$\begin{aligned}\boldsymbol{\beta}_t &= \mathbf{d} + \sum_{n=1}^{p}\mathbf{A}_n\boldsymbol{\beta}_{t-n} + \mathbf{w}_t \quad [\mathbf{w}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{w}),\ t = 1, \ldots, T] \\ &= \mathbf{d} + \mathbf{A}\mathbf{Z}_{t-1} + \mathbf{w}_t, \end{aligned} \tag{5}$$

where $\mathbf{d}$ is a k-dimensional vector, $\mathbf{A}_n$ is a $(k \times k)$ coefficient matrix, $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$ is a $k \times kp$ matrix, $\mathbf{Z}_t = (\boldsymbol{\beta}'_t, \ldots, \boldsymbol{\beta}'_{t-p+1})'$ is a kp-dimensional vector, and $\mathbf{w}_t$ is a k-dimensional white noise term.

Note that a number of common univariate and multivariate stochastic process models, such as random walk, random walk with drift, and AR(p), are special cases of (5). Because $\mathbf{A}_n$ is not assumed to be diagonal, an advantage of the VAR(p) process over the popular dynamic simultaneous equations approach is that it allows one to monitor the relationship between a particular element of $\boldsymbol{\beta}_t$ and a different element of $\boldsymbol{\beta}_{t-n}$. Further, (5) can capture both stationary and nonstationary dynamics; it is well known that the VAR(p) process is stable and thus stationary if

$$\det(\mathbf{I}_k - \mathbf{A}_1 l - \cdots - \mathbf{A}_p l^p) \neq 0 \quad \text{for } |l| \leq 1, \tag{6}$$

that is, if there is no root within or on the unit disk of the reverse characteristic polynomial of the VAR(p) process. If the VAR process is stable, then the expected value of $\boldsymbol{\beta}_t$ does not depend on t, that is,

$$\boldsymbol{\mu}_\beta = E(\boldsymbol{\beta}_t) = (\mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_p)^{-1}\mathbf{d},$$

where the expectation is with respect to $\mathbf{w}_t$, $t = 1, 2, \ldots$ (Lütkepohl 1991).

Finally, we model the heterogeneity of $\mathbf{b}_h$ in (3) as a multivariate normal random effect,

$$p(\mathbf{b}_h|\boldsymbol{\Sigma}_\mathbf{b}) = \mathrm{N}_k(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{b}) \qquad \forall h, \tag{7}$$

where $\boldsymbol{\Sigma}_\mathbf{b}$ is an unknown covariance matrix.

2.3 Prior Distributions

Priors are required for $\{\boldsymbol{\beta}_t\}_{t=1-p}^{0}$, $\mathbf{d}$, $\mathbf{A}$, $\boldsymbol{\Sigma}_\mathbf{w}$, and $\boldsymbol{\Sigma}_\mathbf{b}$. Following standard assumptions of dynamic state-space models (Cargnoni, Müller, and West 1997; Carlin, Polson, and Stoffer 1992; Harrison and Stevens 1976; West and Harrison 1997), we assume that $\{\boldsymbol{\beta}_t\}_{t=1-p}^{0}$, $\mathbf{d}$, $\mathbf{A}$, $\boldsymbol{\Sigma}_\mathbf{w}$, and $\boldsymbol{\Sigma}_\mathbf{b}$ are mutually independent and use the following prior distributions:

$$p(\boldsymbol{\beta}_i) = \mathrm{N}_k(\mathbf{m}_0, \mathbf{S}_0), \quad \text{where } i = 1-p, \ldots, 0; \tag{8}$$

$$p(\mathbf{d}) = \mathrm{N}_k(\mathbf{m}_\mathbf{d}, \mathbf{S}_\mathbf{d}); \tag{9}$$

$$p(\mathrm{vec}(\mathbf{A})) = \mathrm{N}_{k^2p}(\mathbf{m}_\alpha, \mathbf{S}_\alpha); \tag{10}$$

$$p(\boldsymbol{\Sigma}_\mathbf{w}) = \mathrm{IW}_k(\nu_\mathbf{w}, \mathbf{S}_\mathbf{w}); \tag{11}$$

and

$$p(\boldsymbol{\Sigma}_\mathbf{b}) = \mathrm{IW}_k(\nu_\mathbf{b}, \mathbf{S}_\mathbf{b}). \tag{12}$$

Here vec$(\cdot)$ is the usual column stacking operator, so that vec($\mathbf{A}$) is a k^2p-dimensional vector. The expression $p(\boldsymbol{\Sigma}) = \mathrm{IW}_k(\nu, \mathbf{S})$ denotes that $\boldsymbol{\Sigma}$ has a k-dimensional inverted Wishart distribution with parameters ν and $\mathbf{S}$, where $\nu > 0$ and $\mathbf{S}$ is nonsingular, that is, $p(\boldsymbol{\Sigma}) = \mathrm{IW}_k(\nu, \mathbf{S}) \propto |\boldsymbol{\Sigma}|^{-(1/2\nu+k)} \exp(-\frac{1}{2} \times \mathrm{tr}\boldsymbol{\Sigma}^{-1}\mathbf{S})$. Furthermore, the parameters of the prior distributions ($\mathbf{m}_0$, $\mathbf{S}_0$, $\mathbf{m_d}$, $\mathbf{S_d}$, $\mathbf{m}_\alpha$, $\mathbf{S}_\alpha$, $\nu_\mathbf{w}$, $\mathbf{S_w}$, $\nu_\mathbf{b}$, and $\mathbf{S_b}$) are known values that we choose to obtain noninformative proper priors. In particular, for the prior of vec($\mathbf{A}$), we use the Minnesota priors (Doan, Litterman, and Sims 1984; Litterman 1986), which are specialized for the VAR(p) process. Specifically, we choose $\mathbf{m}_\alpha = \mathbf{0}$ and $\mathbf{S}_\alpha$ to be a diagonal matrix with elements

$$s_{rc,n} = \begin{cases} \left(\dfrac{\lambda}{n}\right)^2 & \text{if } r = c \\ \left(\dfrac{\theta\lambda\sigma_r}{n\sigma_c}\right)^2 & \text{otherwise,} \end{cases} \tag{13}$$

where $s_{rc,n}$ is the prior variance of the (r, c) element of $\mathbf{A}_n$ ($n = 1, \dots, p$), σ_r/σ_c is the ratio of square roots of corresponding diagonal elements of $\boldsymbol{\Sigma}_\mathbf{w}$, λ is the prior belief on the tightness around 0 for the diagonal elements of $\mathbf{A}_1$, and $0 < \theta < 1$ reflects the prior belief that most of the variation of $\boldsymbol{\beta}_t$ is explained by its own lags. Thus the Minnesota priors are locally noninformative proper priors around 0, an attractive property because, under the stability condition, $\mathbf{A}_n$ tends to shrink to 0 rapidly in n (cf. Lütkepohl 1991, p. 208). The Minnesota priors can also be characterized as smoothly decreasing priors over lags in a harmonic manner, which is also useful for order selection of p.

3. ESTIMATION AND MODEL CHOICE

We first discuss parameter estimation of the proposed dynamic logit model, then describe the model selection procedure.

3.1 Full Posterior Distribution

Using the likelihood and prior specifications, we obtain the posterior distribution for all parameters. Let

- $\mathcal{H} = \{1, 2, \dots, H\}$ be the set of all individuals,
- $\mathcal{H}_t$ be a subset of $\mathcal{H}$ that consists of individuals that make choices at time t,
- $\mathbf{y}_t = \{y_{ht}\}_{h\in\mathcal{H}_t}$ denote the observed choice data at time t,
- $\mathbf{y} = (\mathbf{y}_1', \dots, \mathbf{y}_T')'$ denote all choice data from time 1 to time T,

and

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \dots, \boldsymbol{\beta}_T')'$, $\mathbf{b}_t = \{\mathbf{b}_h\}_{h\in\mathcal{H}_t}$, and $\mathbf{b} = \{\mathbf{b}_h\}_{h\in\mathcal{H}}$.

Then the posterior distribution is

$$\begin{aligned} p(&\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}|\mathbf{y}) \\ &\propto \left(\prod_{t=1}^{T} p(\mathbf{y}_t|\boldsymbol{\beta}_t, \mathbf{b}_t)\right) \times \left(\prod_{n=1-p}^{0} p(\boldsymbol{\beta}_n)\right) \\ &\quad \times \left(\prod_{t=1}^{T} p(\boldsymbol{\beta}_t|\mathbf{d}, \mathbf{A}, \boldsymbol{\beta}_{t-1}, \dots, \boldsymbol{\beta}_{t-p}, \boldsymbol{\Sigma}_\mathbf{w})\right) \\ &\quad \times \left(\prod_{\mathcal{H}} p(\mathbf{b}_h|\boldsymbol{\Sigma}_\mathbf{b})\right) \times p(\mathbf{d}) \times p(\mathbf{A}) \\ &\quad \times p(\boldsymbol{\Sigma}_\mathbf{w}) \times p(\boldsymbol{\Sigma}_\mathbf{b}), \end{aligned} \tag{14}$$

where

$$p(\mathbf{y}_t|\boldsymbol{\beta}_t, \mathbf{b}_t) = \prod_{h\in\mathcal{H}_t} \prod_{j=1}^{J} p_{hjt}^{q_{hjt}},$$

with $q_{hjt} = 1$ if $y_{ht} = j$ and $q_{hjt} = 0$ otherwise.

This posterior distribution has several sets of parameters, with numerous elements. Specifically, for $\boldsymbol{\beta}_t$ ($t = 1-p, \dots, T$), there are $k(p+T)$; for $\mathbf{b}_h$ ($h = 1, \dots, H$), kH; for $\mathbf{d}$, k; for $\mathbf{A}$, k^2p; and for both $\boldsymbol{\Sigma}_\mathbf{w}$ and $\boldsymbol{\Sigma}_\mathbf{b}$, $\frac{1}{2}k(k+1)$. In the forthcoming illustration, we have $H = 492$, $T = 90$, and $k = 6$, yielding $3{,}540 + 42p$ elements overall.

Because analytic methods to evaluate the posterior distribution in (14) are not available, we use Markov chain Monte Carlo (MCMC) methods, as described in Section 3.2. For model selection and Bayesian hypothesis testing, we use Bayes factors (Bernardo and Smith 1994) to compare two models (or hypotheses) H_1 and H_2,

$$B_{12} = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_2)} = \frac{\int p(\mathbf{y}|\Phi, H_1)p(\Phi|H_1)\, d\Phi}{\int p(\mathbf{y}|\Psi, H_2)p(\Psi|H_2)\, d\Psi}, \tag{15}$$

where $p(\cdot|H_i)$ is the prior on parameters under model i ($i = 1, 2$) and $p(\mathbf{y}|\cdot, H_i)$ is the likelihood under model i. To estimate the Bayes factor, we must evaluate the integrated likelihoods, using the results from the MCMC simulation. Let $\Phi^{(g)}$, $g = 1, \dots, G$, denote the G values of Φ generated from the posterior distribution of Φ, $p(\Phi|H_1)$. The integrated likelihood for model 1, $p(\mathbf{y}|H_1) = \int p(\mathbf{y}|\Phi, H_1)p(\Phi|H_1)\, d\Phi$ in (15), can be estimated by the harmonic mean estimator (Newton and Raftery 1994),

$$\widehat{p}(\mathbf{y}|H_1) = \left(\frac{1}{G}\sum_{g=1}^{G} \frac{1}{p(\mathbf{y}|\Phi^{(g)}, H_1)}\right)^{-1}.$$

This estimator converges almost surely to the correct value, but it does not generally satisfy a Gaussian central limit theorem. Nevertheless, it has been found to work reasonably well with large samples (cf. Kass and Raftery 1995).

3.2 Markov Chain Monte Carlo Sampler

To evaluate the posterior distribution, $p(\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}|\mathbf{y})$, given in (14), we implement a MCMC sampler, using the following conditional posterior distributions:

$$\begin{aligned} p(&\boldsymbol{\beta}|\mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y}) \\ &\leftrightarrow p(\mathbf{b}|\boldsymbol{\beta}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y}) \leftrightarrow p(\mathbf{d}|\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y}) \\ &\leftrightarrow p(\mathbf{A}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y}) \leftrightarrow p(\boldsymbol{\Sigma}_\mathbf{w}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y}) \\ &\leftrightarrow p(\boldsymbol{\Sigma}_\mathbf{b}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \mathbf{y}). \end{aligned}$$

We next describe the sampling procedure for each of these.

3.2.1 Sampling From $p(\boldsymbol{\beta}|\mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y})$. To sample from $p(\boldsymbol{\beta}|\mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y})$, we need the conditional posterior density for $\boldsymbol{\beta}_t$, $p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{m\neq t}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y})$. When $1 \leq t \leq T$,

$$
\begin{aligned}
p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{m\neq t}, &\mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y}) \\
&\propto \prod_{m=0}^{p} p(\boldsymbol{\beta}_{t+m}|\mathbf{d}, \mathbf{A}, \boldsymbol{\beta}_{t+m-1}, \ldots, \boldsymbol{\beta}_{t+m-p}, \boldsymbol{\Sigma}_{\mathbf{w}}) \\
&\quad \times p(\mathbf{y}_t|\boldsymbol{\beta}_t, \mathbf{b}_t) \qquad (16) \\
&= \mathrm{N}(\mathbf{F}_t\mathbf{f}_t, \mathbf{F}_t)p(\mathbf{y}_t|\boldsymbol{\beta}_t, \mathbf{b}_t), \qquad (17)
\end{aligned}
$$

where

$$
\mathbf{F}_t^{-1} = \begin{cases}
\mathbf{S}_0^{-1} + \sum_{m=1-t}^{p} \mathbf{A}_m' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}_m, & t = 1-p, \ldots, -1, 0 \\
\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \sum_{m=1}^{p} \mathbf{A}_m' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}_m, & t = 1, \ldots, T-p \\
\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \sum_{m=1}^{T-t} \mathbf{A}_m' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}_m, & t = T-p+1, \ldots, T-1 \\
\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}, & t = T,
\end{cases}
$$

and

$$
f_t' = \begin{cases}
\mathbf{m}_0'\mathbf{S}_0^{-1} + \sum_{m=1-t}^{p} \left(\boldsymbol{\beta}_{t+m} - \mathbf{d} - \sum_{n=1, n\neq m}^{p} \mathbf{A}_n \boldsymbol{\beta}_{t+m-n}\right)' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}_m', & t = 1-p, \ldots, -1, 0 \\
\left(\mathbf{d} + \sum_{m=1}^{p} \mathbf{A}_m \boldsymbol{\beta}_{t-m}\right)' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \sum_{m=1}^{p} \left(\boldsymbol{\beta}_{t+m} - \mathbf{d} - \sum_{n=1, n\neq m}^{p} \mathbf{A}_n \boldsymbol{\beta}_{t+m-n}\right)' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}_m', & t = 1, \ldots, T-p \\
\left(\mathbf{d} + \sum_{m=1}^{p} \mathbf{A}_m \boldsymbol{\beta}_{t-m}\right)' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \sum_{m=1}^{T-t} \left(\boldsymbol{\beta}_{t+m} - \mathbf{d} - \sum_{n=1, n\neq m}^{p} \mathbf{A}_n \boldsymbol{\beta}_{t+m-n}\right)' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}_m', & t = T-p+1, \ldots, T-1 \\
\left(\mathbf{d} + \sum_{m=1}^{p} \mathbf{A}_m \boldsymbol{\beta}_{T-m}\right)' \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}, & t = T.
\end{cases}
$$

Given (17), a Metropolis–Hastings algorithm step can be conducted as follows (Chib and Greenberg 1995; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953):

1. Sample $\boldsymbol{\beta}_t^*$ from a proposal density, $\mathrm{N}(\boldsymbol{\beta}_t^{\mathrm{pre}}, \phi_{\beta}\mathbf{I})$, where $\boldsymbol{\beta}_t^{\mathrm{pre}}$ is the most recently updated value and ϕ_{β} is a fixed tuning constant.
2. Substitute $\boldsymbol{\beta}_t^*$ for $\boldsymbol{\beta}_t^{\mathrm{pre}}$ with acceptance probability

$$
\pi(\boldsymbol{\beta}_t^*, \boldsymbol{\beta}_t^{\mathrm{pre}}) = \min\left(\frac{p(\mathbf{y}_t|\boldsymbol{\beta}_t^*, \mathbf{b}_t)N(\boldsymbol{\beta}_t^*|\mathbf{F}_t\mathbf{f}_t, \mathbf{F}_t)}{p(\mathbf{y}_t|\boldsymbol{\beta}_t^{\mathrm{pre}}, \mathbf{b}_t)N(\boldsymbol{\beta}_t^{\mathrm{pre}}|\mathbf{F}_t\mathbf{f}_t, \mathbf{F}_t)}, 1\right),
$$

where $\mathrm{N}(\boldsymbol{\beta}_t|\mathbf{F}_t\mathbf{f}_t, \mathbf{F}_t)$ denotes the multivariate normal density with mean $\mathbf{F}_t\mathbf{f}_t$ and covariance matrix $\mathbf{F}_t$, evaluated at $\boldsymbol{\beta}_t$.

3.2.2 Sampling From $p(\mathbf{b}|\boldsymbol{\beta}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y})$ and $p(\mathbf{d}|\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y})$. The conditional posterior density for $\mathbf{b}_h$ is

$$
p(\mathbf{b}_h|\boldsymbol{\beta}_t, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, y_{ht}) \propto p(\mathbf{b}_h|\boldsymbol{\Sigma}_{\mathbf{b}}) \prod_{t_h} p(y_{ht}|\boldsymbol{\beta}_t, \mathbf{b}_h),
$$

where $t_h = \{t : h \in \mathcal{H}_t\}$. A Metropolis–Hastings step can be used, as follows:

1. Sample $\mathbf{b}_h^*$ from a proposal density, $N(\mathbf{b}_h^{\mathrm{pre}}, \phi_{\mathbf{b}}\mathbf{I}_k)$, where $\mathbf{b}_h^{\mathrm{pre}}$ is the most recently updated value and $\phi_{\mathbf{b}}$ is a fixed tuning constant.
2. Substitute $\mathbf{b}_h^*$ for $\mathbf{b}_h^{\mathrm{pre}}$ with acceptance probability

$$
\pi(\mathbf{b}_h^*, \mathbf{b}_h^{\mathrm{pre}}) = \min\left(\frac{p(\mathbf{b}_h^*|\boldsymbol{\Sigma}_{\mathbf{b}})\prod_{t_h} p(y_{ht}|\boldsymbol{\beta}_t, \mathbf{b}_h^*)}{p(\mathbf{b}_h^{\mathrm{pre}}|\boldsymbol{\Sigma}_{\mathbf{b}})\prod_{t_h} p(y_{ht}|\boldsymbol{\beta}_t, \mathbf{b}_h^{\mathrm{pre}})}, 1\right).
$$

The conditional posterior density for $\mathbf{d}$ is

$$
\begin{aligned}
p(\mathbf{d}|\boldsymbol{\beta}, \mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y}) &\propto \prod_{t=1}^{T} p(\boldsymbol{\beta}_t|\mathbf{d}, \mathbf{A}, \mathbf{Z}_{t-1}, \boldsymbol{\Sigma}_{\mathbf{w}})p(\mathbf{d}) \\
&= \mathrm{N}(\boldsymbol{\mu}_{\mathbf{d}}^*, \boldsymbol{\Sigma}_{\mathbf{d}}^*), \qquad (18)
\end{aligned}
$$

a multivariate normal density with mean vector $\boldsymbol{\mu}_{\mathbf{d}}^* = \boldsymbol{\Sigma}_{\mathbf{d}}^* \times \{\mathbf{S}_{\mathbf{d}}^{-1}\mathbf{m}_{\mathbf{d}} + \sum_{t=1}^{T} \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\boldsymbol{\beta}_t - \mathbf{A}\mathbf{Z}_{t-1})\}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{d}}^* = (\mathbf{S}_{\mathbf{d}}^{-1} + T\boldsymbol{\Sigma}_{\mathbf{w}}^{-1})^{-1}$.

3.2.3 Sampling From $p(\mathbf{A}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y})$. Let the (k^2p)-vector $\boldsymbol{\alpha} = \mathrm{vec}(\mathbf{A})$; let $\mathbf{Z} = (Z_0, Z_1, \ldots, Z_{T-1})$, $\tilde{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t - \mathbf{d}$, and $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_T)$; and let the kT-vector $\tilde{\boldsymbol{\beta}} = \mathrm{vec}(\tilde{\boldsymbol{\beta}})$. Then the conditional posterior of $\boldsymbol{\alpha}$ is

$$
\begin{aligned}
p(\boldsymbol{\alpha}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\mathbf{w}}) &= p(\boldsymbol{\alpha}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{b}}, \mathbf{y}) \propto p(\tilde{\boldsymbol{\beta}}|\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\mathbf{w}})p(\boldsymbol{\alpha}) \\
&\propto \exp\Big[-\frac{1}{2}\{(\tilde{\boldsymbol{\beta}} - (\mathbf{Z}' \otimes \mathbf{I}_k)\boldsymbol{\alpha})'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})(\tilde{\boldsymbol{\beta}} - (\mathbf{Z}' \otimes \mathbf{I}_k)\boldsymbol{\alpha}) \\
&\qquad + (\boldsymbol{\alpha} - \mathbf{m}_{\alpha})'\mathbf{S}_{\alpha}^{-1}(\boldsymbol{\alpha} - \mathbf{m}_{\alpha})\}\Big].
\end{aligned}
$$

By completing the square in $\boldsymbol{\alpha}$,

$$
p(\boldsymbol{\alpha}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\mathbf{w}}) = \mathrm{N}(\boldsymbol{\alpha}^*, \boldsymbol{\Sigma}_{\alpha}^*), \qquad (19)
$$

a (k^2p)-dimension normal density with mean vector $\boldsymbol{\alpha}^* = \boldsymbol{\Sigma}_{\alpha}^*\{\mathbf{S}_{\alpha}^{-1}\mathbf{m}_{\alpha} + (\mathbf{Z} \otimes \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\tilde{\boldsymbol{\beta}}\}$ and covariance matrix $\boldsymbol{\Sigma}_{\alpha}^* = [\mathbf{S}_{\alpha}^{-1} + (\mathbf{Z}\mathbf{Z}' \otimes \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})]^{-1}$.

If we do not impose the stability restriction on the VAR(p) process, then (19) can be used directly to sample $\boldsymbol{\alpha}$. In this case the probability of the VAR(p) process being stable can be estimated by counting the number of iterations when the sampled $\boldsymbol{\alpha}$ satisfies (6). However, with a stability restriction on the VAR(p) process, there is a difficulty in sampling $\boldsymbol{\alpha}$. Under the stability condition (6), $\boldsymbol{\alpha}$ should be sampled from $\mathrm{N}(\boldsymbol{\alpha}^*, \boldsymbol{\Sigma}_{\alpha}^*)I(\boldsymbol{\alpha} \in B)$, where B is the region in which the stability condition is satisfied. The simplest way to sample $\boldsymbol{\alpha}$ under the stability restriction is to use rejection sampling, by accepting $\boldsymbol{\alpha}$ sampled from $\mathrm{N}(\boldsymbol{\alpha}^*, \boldsymbol{\Sigma}_{\alpha}^*)$ only if it satisfies (6). However, rejection sampling will be inefficient, because the rejection rate increases exponentially with the dimension of $\boldsymbol{\alpha}$. Even for a univariate AR(p) process, the acceptance rate of rejection sampling approaches 50% (e.g., Barnett, Kohn, and Sheather 1996).

We therefore sample $\boldsymbol{\alpha}$ directly from $\mathrm{N}(\boldsymbol{\alpha}^*, \boldsymbol{\Sigma}_\alpha^*)I(\boldsymbol{\alpha} \in B)$ under the stability restriction by using single-variable slice-sampling, as proposed by Neal (1997). Recall that $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{k^2p})'$, and consider the conditional distribution of α_i, $f(\alpha_i) = p(\alpha_i|\text{remaining components of } \alpha_i)$, which is proportional to $\mathrm{N}(\boldsymbol{\alpha}^*, \boldsymbol{\Sigma}_\alpha^*)I(\boldsymbol{\alpha} \in B)$. Generating values of α_i proceeds by replacing the previous value, α_i^{pre}, with a new value, α_i^{new}, as follows:

1. Define a horizontal slice, $S^h = \{\alpha_i : z < f(\alpha_i)\}$, where z is an auxiliary variable sampled uniformly from $(0, f(\alpha_i^{\text{pre}}))$.
2. Find an interval, $I = (L, R)$, around α_i^{pre} on S^h such that $f(L) < z$ and $f(R) < z$.
3. Accept α_i^{new}, sampled uniformly from I, if $f(\alpha_i^{\text{new}}) > z$.

Roberts and Rosenthal (1999) showed that the slice sampler is irreducible and aperiodic and satisfies the detailed-balance condition. The advantages of the slice sampler are that it can be used for any log-concave probability density function. Furthermore, it can avoid slow random-walk convergence, because α_i^{pre} is always replaced by α_i^{new} in each iteration and it is possible to obtain a large jump from α_i^{pre} to α_i^{new}. Computation of $f(\alpha_i)$ involves the evaluation of $I(\boldsymbol{\alpha} \in B)$. Note that the evaluation of $I(\boldsymbol{\alpha} \in B)$ does not require computation of lower and upper bounds of the region B. One must simply check whether or not α_i falls inside the region B, by using (6).

In some cases, a researcher may have prior beliefs on $\boldsymbol{\alpha}$ (or, equivalently, A) and so wishes to place restrictions on a subset of $\boldsymbol{\alpha}$. In such a case, $\boldsymbol{\alpha}$, under arbitrary restrictions, can be easily sampled as follows. Suppose that $\bar{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}_1'; \boldsymbol{\alpha}_2')'$ with $\boldsymbol{\alpha}_2 = \mathbf{a}$, where $\mathbf{a}$ is a vector of restricted values. Define a partition matrix $\mathbf{P}$ such that $\bar{\boldsymbol{\alpha}} = \mathbf{P}\boldsymbol{\alpha}$. Then the conditional posterior density of $\boldsymbol{\alpha}_1$ given $\boldsymbol{\alpha}_2 = \mathbf{a}$, $\mathrm{N}(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_2 = \mathbf{a})$, can be easily obtained from $\mathrm{N}(\mathbf{P}\boldsymbol{\alpha}^*, \mathbf{P}\boldsymbol{\Sigma}_\alpha^*\mathbf{P}')$. If the partitioned submatrix for $\boldsymbol{\alpha}_2$ in $\mathbf{P}\boldsymbol{\Sigma}_\alpha^*\mathbf{P}'$ is singular, the Moore–Penrose inverse can be used to derive $\mathrm{N}(\boldsymbol{\alpha}_1|\boldsymbol{\alpha}_2 = \mathbf{a})$.

3.2.4 Sampling From $p(\boldsymbol{\Sigma}_\mathbf{w}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y})$ and $p(\boldsymbol{\Sigma}_\mathbf{b}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \mathbf{y})$. The conditional posterior density for $\boldsymbol{\Sigma}_\mathbf{w}$ is

$$p(\boldsymbol{\Sigma}_\mathbf{w}|\boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{b}, \mathbf{y}) = p(\boldsymbol{\Sigma}_\mathbf{w}|\boldsymbol{\beta}, \{\boldsymbol{\beta}\}_{m=1-p}^0, \mathbf{d}, \mathbf{A}, \mathbf{y}) \propto \mathrm{IW}(\nu_\mathbf{w}^*, \mathbf{S}_\mathbf{w}^*),$$

an inverted Wishart density with $\nu_\mathbf{w}^* = \nu_\mathbf{w} + T$ and $\mathbf{S}_\mathbf{w}^* = \mathbf{S}_\mathbf{w} + \sum_{t=1}^T \mathbf{l}_t\mathbf{l}_t'$, where $\mathbf{l}_t = \boldsymbol{\beta}_t - \mathbf{d} - \mathbf{A}\mathbf{Z}_{t-1}$.

The conditional posterior density for $\boldsymbol{\Sigma}_\mathbf{b}$ is

$$p(\boldsymbol{\Sigma}_\mathbf{b}|\mathbf{b}) = p(\boldsymbol{\Sigma}_\mathbf{b}|\mathbf{y}, \boldsymbol{\beta}, \mathbf{b}, \mathbf{d}, \mathbf{A}, \boldsymbol{\Sigma}_\mathbf{w}, \mathbf{y}) \propto \mathrm{IW}(\nu_\mathbf{b}^*, \mathbf{S}_\mathbf{b}^*), \tag{20}$$

an inverted Wishart density with $\nu_\mathbf{b}^* = \nu_\mathbf{b} + H$ and $\mathbf{S}_\mathbf{b}^* = \mathbf{S}_\mathbf{b} + \sum_{h \in \mathcal{H}} \mathbf{b}_h\mathbf{b}_h'$.

3.3 Comparative Model Specifications

The model (2) has parameters $\{\boldsymbol{\beta}_t\}_{t=1}^T$, $\{\mathbf{b}_h\}_{h=1}^H$, $\mathbf{d}$, $\{\mathbf{A}_n\}_{n=1}^p$, $\boldsymbol{\Sigma}_\mathbf{w}$, and $\boldsymbol{\Sigma}_\mathbf{b}$. We consider several alternative models that differ by the structural assumptions imposed on $\mathbf{d}$, $\{\mathbf{A}_n\}_{n=1}^p$, and $\boldsymbol{\Sigma}_\mathbf{w}$, as follows:

Model	$\mathbf{d}$	$\{\mathbf{A}_n\}_{n=1}^p$	$\boldsymbol{\Sigma}_\mathbf{w}$
No parameter dynamics			
M_0: *Static random-effects logit model*	NR	**0**	**0**
Parameter dynamics			
M_1: Dynamic linear model; random walk	**0**	$p=1$; $\mathbf{A}_1 = \mathbf{I}$	NR
M_2: Random walk with a drift	NR	$p=1$; $\mathbf{A}_1 = \mathbf{I}$	NR
M_3: VAR(p)	NR	NR	NR
M_4: Restricted VAR(p); RVAR(p)	NR	$\mathbf{A}_n =$ diagonal	NR

NOTE: *NR, no restriction.*

All models listed in this table incorporate heterogeneity as a random-effects specification; see (7). Model M_0 is the traditional random-effect logit, which assumes that there are no parameter dynamics. Models M_1–M_4 allow for parameter dynamics in different ways. Model M_1, the popular dynamic linear model (Harrison and Stevens 1976; West and Harrison 1997), assumes a random-walk process for $\boldsymbol{\beta}_t$ with mean vector $\boldsymbol{\beta}_{t-1}$ and covariance matrix $\boldsymbol{\Sigma}_\mathbf{w}$. Model M_2 assumes a random-walk process with a drift term for $\boldsymbol{\beta}_t$. Model M_3 is the proposed VAR(p) process model. Model M_4 is a restricted VAR(p) [RVAR(p)] process model under the restriction that $\{\mathbf{A}_n\}$ are diagonal matrices. Thus, in terms of parametric restriction, $M_1 \subset M_2 \subset M_4 \subset M_3$.

M_0 can be easily estimated by skipping the MCMC steps for $\boldsymbol{\beta}_t$, $\boldsymbol{\Sigma}_\mathbf{w}$, and vec($\mathbf{A}$). M_1 can be estimated by skipping the MCMC steps of $\mathbf{d}$ and vec($\mathbf{A}$). Similarly, M_2 can be estimated by skipping the MCMC step for vec($\mathbf{A}$). We estimate M_3 and M_4 under the stability condition given in (6).

To test the accuracy of parameter recovery, we performed two extensive simulation studies that differed in the relative complexity of $\boldsymbol{\beta}_t$'s dynamics. All model parameters were recovered well in each. (Full results are available from the authors.)

4. EMPIRICAL ILLUSTRATION

4.1 Data and Independent Variables

The proposed model was estimated on A. C. Nielsen liquid detergent scanner data over 96 weeks. The data consist of 492 individual units (households) that made choices among four options $\{A, B, C, D\}$ at least seven times during the 96-week period. The first 90 weeks of data were used as a training sample to estimate the model, and the remaining 6 weeks of data were used for predicting the future model parameters. The numbers of observations for the training and future parameter forecasting samples were 6,364 and 318. To ensure identifiability, the time-varying common effect of the fourth option as well as its random effect were fixed to be 0. This requires that the values $\mathbf{x}_{hjt}$ for option j be the *differences* of the corresponding predictor variable values for options j and the base option, J. The vector $\mathbf{x}_{hjt}$ thus consists of three option dummies and three covariates, the differences in feature, display and price; note that the first two are binary, whereas the last one is continuous.

All mean vectors of the normal priors [i.e., $\mathbf{m}_0$, $\mathbf{m}_\mathbf{d}$, and $\mathbf{m}_\alpha$ in (8) to (10)] were set to 0. The chosen values for $\mathbf{S}_0$ and $\mathbf{S}_\mathbf{d}$ were 100$\mathbf{I}$. For the inverse-Wishart priors of $\boldsymbol{\Sigma}_\mathbf{w}$ and $\boldsymbol{\Sigma}_\mathbf{b}$, the degree of freedom parameters were set at 2 and the scale parameters were chosen so as to make the expected values 100**I**. For the value of $\mathbf{S}_\alpha$ in (10) and (13), Litterman (1986) suggested choosing λ, θ, and $\{\sigma_i\}$ by examining the data and trying several different values. However, this approach entails double usage of data. We thus set $\lambda = 1.0$, θ to .5, and all ratios σ_r/σ_c to 1. For VAR(p) when $p > 1$, we set $\mathbf{S}_\alpha$ using (13) with $\lambda = 1.0$, $\theta = .5$, and $\frac{\sigma_r}{\sigma_c} = 1$.

4.2 Markov Chain Monte Carlo Estimation

The tuning constants for the proposal distributions in the Metropolis–Hastings algorithms (e.g., ϕ_β in Sec. 3.2.1) were chosen to produce similar acceptance rates across models. There exists a trade-off between convergence speed and acceptance rate in the Metropolis–Hastings algorithm (Chib and Greenberg 1995). As tuning constants become smaller, the acceptance rate increases, but we need a longer chain, because the distance between the previous value and a newly accepted value becomes smaller. The chosen tuning constants for $\boldsymbol{\beta}_t$ and $\mathbf{b}_h$ were approximately $\phi_\beta = .07$ and $\phi_{\mathbf{b}} = .3$. For all models, the acceptance rates for $\boldsymbol{\beta}_t$ and $\mathbf{b}$, given these tuning constants, ranged from 53.2% to 55.7% and from 53.8% to 57.5%.

The number of quantities of interest for the VAR(p) model is quite large. For example, excluding $\mathbf{b}$, there are 630 quantities for the full VAR(1) model. Thus we must be careful in determining convergence of the MCMC sampler. Specifically, we determine convergence after examining all quantities except $\mathbf{b}$. To monitor convergence, we use Geweke's (1992) convergence diagnostic, which is based on the smooth spectral density of a MCMC posterior sample. The periodogram for spectral density estimation involves two important choices: window and truncation point. We use the Tukey window and choose the truncation point after looking at the autocovariance function, as Jenkins and Watts (1968) suggested.

After 20,000 iterations, all models seem to reach convergence. Figure 1 is a typical example; for the VAR(1) model, it shows the trace plot for the six elements of $\boldsymbol{\beta}_1$ for the first 40,000 iterations, where G represents Geweke's convergence diagnostic with 20,000 burn-in periods. Across all models, the proportion of quantities that pass the Geweke diagnostic ranges from 91.3% to 98.1%. For the rejected quantities, we used Heidelberger and Welch's (1983) half-width test, which the majority passed. All inferences made here are based on the next 20,000 iterations.

4.3 Tests for Parameter Dynamics

After estimating models M_0–M_4, we can test whether there is evidence that parameters are time-varying. Specifically, we have the following hypotheses:

H_1 : Parameters are static (M_0)

and

H_2 : Parameters display some form of dynamics (M_1, M_2, M_3, or M_4).

The computed integrated likelihoods and the Bayes factors for a comparison of models M_0 and M_i (BF_{M_0,M_i}) are given in Table 1. Because VAR(2) has a smaller integrated likelihood than VAR(1) and its estimated A_2 is close to a null matrix, we do not estimate VAR(p) models of higher orders. We also do not estimate RVAR(p) with order greater than 2, because RVAR(2) shows smaller integrated likelihood than RVAR(1), and its estimated A_2 is close to null.

As shown in Table 1, we find exceptionally strong evidence supporting parameter dynamics. All models incorporating parameter dynamics (M_1–M_4) are decisively preferred over the traditional static random-effects model (M_0). Therefore, the model parameters, taken as a set, are evidently time-varying.

Selection Among Dynamic Models. The most preferred among the parameter dynamics models (M_1–M_4) is RVAR(1), as shown in Table 1. The Bayes factors for RVAR(1) against the other dynamic models range from $9.8e{+}9$ to $9.3e{+}15$. Most interestingly, the RVAR(1) model is decisively preferred to the full VAR(1) model (Bayes factor $= .98e{+}10$), suggesting that A_1 is diagonal or very nearly so. Further, the VAR(1) and RVAR(1) models are preferred over M_1 and M_2, implying that the matrix A_1 is not an identity matrix; furthermore, the value of A_1, reported later, suggests stable parameter dynamics.

4.4 Cross-Validation

One can appeal to cross-validation to compare M_0, the static random-effects model, with the RVAR(1) model. To do this, we divide the 96 weeks of data into two sets. The "calibration data set" consists of the first w weeks of data and is used for parameter estimation, whereas the "prediction dataset" consists of the remaining $96 - w$ weeks and is used, unsurprisingly, for prediction. We investigated values for $w = \{50, 55, 60, 65, 70\}$ to assess how additional calibration data affects predictive accuracy. For the calibration dataset, we computed the Bayes factor of M_0 versus RVAR(1). Results are given in Table 2. Regardless of the value for w, model RVAR(1) is decisively preferred to M_0.

The results for the prediction dataset were based on the following approach. For both models M_0 and RVAR(1), we estimated the likelihood for the prediction data set as follows. For model M_0, we estimated this likelihood by first computing choice probabilities (2) for the prediction dataset (given $\mathbf{b}_h$ and $\boldsymbol{\beta}$ simulated at each MCMC iteration) and taking an average of these choice probabilities for the prediction dataset across MCMC iterations. For model RVAR(1), we did the same given each value of $\mathbf{b}_h$ and $\boldsymbol{\beta}_t$, $t = w, \ldots, 96$.

The resulting estimated log-likelihoods are also given in Table 2. The RVAR(1) model is moderately preferred to model M_0, so we next investigate RVAR(1) in more detail.

4.5 Estimation for the Training Sample

Because the RVAR(1) model performs better than other VAR(p) models (see Table 1), we report further estimation results for this model alone. The RVAR(1) model implies the following structure for the regression parameter vector $\boldsymbol{\beta}_{ht}$:

$$\boldsymbol{\beta}_{ht} = \boldsymbol{\beta}_t + \mathbf{b}_h,$$

$$\boldsymbol{\beta}_t = \mathbf{d} + \mathbf{A}_1\boldsymbol{\beta}_{t-1} + \mathbf{w}_t,$$

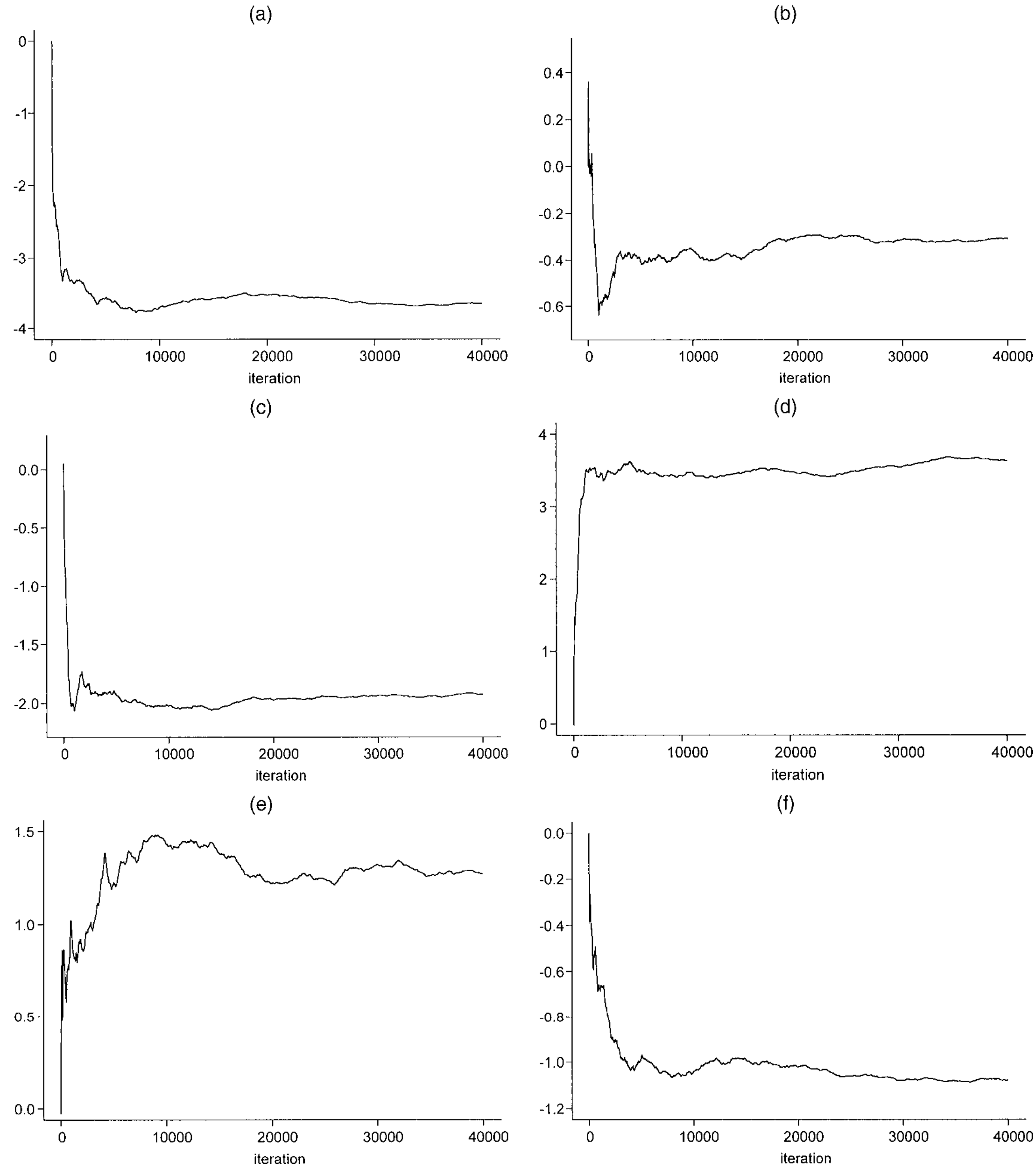

Figure 1. Running Mean Plot of β_1 for the Full VAR(1) Model. (a) dum_A ($G = -.23$); (b) dum_B ($G = 1.15$); (c) dum_C ($G = .34$); (d) feature ($G = -1.54$); (e) display ($G = 1.17$); (f) price ($G = -.59$). G denotes Geweke's convergence diagnostic.

Table 1. Model Comparison for Training Sample

M_i	Log of integrated likelihood	Bayes factor (BF_{M_0,M_i})
No parameter dynamics case		
M_0: Static random-effects logit model	−3,436.33	1.0
Parameter dynamics case		
M_1: Dynamic linear model	−3,068.50	1.79e−160
M_2: Random walk with a drift	−3,072.44	9.22e−159
M_3: VAR(p)		
VAR(1)	−3,061.90	2.44e−163
VAR(2)	−3,075.66	2.31e−157
M_4: RVAR(p)		
RVAR(1)	−3,038.89	2.48e−173
RVAR(2)	−3,063.15	8.51e−163

and

$$\mathbf{b}_h \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{b}), \qquad \mathbf{w}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{w}),$$

where $\mathbf{A}_1$ is a diagonal matrix. The parameters, apart from $\boldsymbol{\beta}_t$ and $\mathbf{b}_h$, are thus $\mathbf{d}$, diag($\mathbf{A}_1$), $\boldsymbol{\Sigma}_\mathbf{w}$, and $\boldsymbol{\Sigma}_\mathbf{b}$.

An important derived parameter is the long-run mean of $\boldsymbol{\beta}_t$, $\boldsymbol{\mu}_\beta = (\mathbf{I} - \mathbf{A}_1)^{-1}\mathbf{d}$. Furthermore, the long-run variance for the ith element of $\boldsymbol{\beta}_t$ is

$$\sigma^2_{\beta,i} = \frac{\Sigma_{\mathbf{w},i,i}}{1 - A^2_{1,i,i}}, \tag{21}$$

where $\Sigma_{\mathbf{w},i,i}$ and $A_{1,i,i}$ are the ith diagonal elements of $\boldsymbol{\Sigma}_\mathbf{w}$ and $\mathbf{A}_1$. (See Hamilton 1994 for derivation of the moments of the full VAR(p) process.) An indication of the overall variabil-

Table 2. Cross-Validation Comparisons

	Calibration sample			Log-likelihood for prediction sample	
	log(integrated likelihood)		Bayes factor		
Data used	M_0	RVAR(1)	(H1: M_0)	M_0	RVAR(1)
Weeks 1–50	−2,048.60	−1,782.73	3.42e−116	−2,256.17	−2,236.09
Weeks 1–55	−2,247.86	−1,946.93	1.33e−123	−1,942.92	−1,916.08
Weeks 1–60	−2,449.33	−2,166.87	2.13e−123	−1,623.90	−1,613.33
Weeks 1–65	−2,571.30	−2,315.47	7.84e−112	−1,396.55	−1,388.50
Weeks 1–70	−2,812.06	−2,516.53	4.50e−129	−1,080.21	−1,071.03

ity for element i of $\boldsymbol{\beta}_{ht}$ is thus given by

$$\mathrm{var}(\beta_{ht,i}) = \frac{\Sigma_{\mathbf{w},i,i}}{1 - A^2_{1,i,i}} + \Sigma_{\mathbf{b},i,i}, \tag{22}$$

which can be used to give an idea of the relative contributions of parameter dynamics and heterogeneity. It can also be compared with the elements of $E(\boldsymbol{\beta}_{ht}) = \boldsymbol{\mu}_\beta = (\mathbf{I} - \mathbf{A}_1)^{-1}\mathbf{d}$ to get an idea of the relative variability of the elements of $\boldsymbol{\beta}_{ht}$.

Estimation of $\boldsymbol{\beta}_t$. Figure 2 plots posterior means and the 5th and 95th percentiles for all time-varying parameters $\boldsymbol{\beta}_t$. It suggests fairly large temporal fluctuations; all intercepts show strong stochastic patterns, and all variable coefficients display stochastic dynamics. There are several periods that show substantial shifts from $\boldsymbol{\beta}_{t-1}$ to $\boldsymbol{\beta}_t$.

Estimation of $\mathbf{d}$ *and* $\mathbf{A}_1$. If we define "significant difference" to mean that a (5th percentile, 95th percentile) interval does not contain 0, Table 3 suggests that all elements of $\mathbf{d}$, except the coefficients of the dummy for option C (dum_C) and of feature, are significantly different from 0. Likewise, the elements of $\mathbf{A}_1$ corresponding to dum_A, dum_B, and Price are significantly different from 0, which suggests systematic dynamics over time for the corresponding elements of $\boldsymbol{\beta}_t$.

Estimation of $\boldsymbol{\Sigma}_\mathbf{w}$. Table 4 gives estimates for $\boldsymbol{\Sigma}_\mathbf{w}$. Posterior means are given on and below the main diagonal, with posterior standard deviations given in parentheses; the posterior means of the correlation coefficients are given above the main diagonal. For the diagonal elements of $\boldsymbol{\Sigma}_\mathbf{w}$, the ratios of posterior means to posterior standard deviations are between 4.8 and 6.1. Thus, these elements of $\boldsymbol{\Sigma}_\mathbf{w}$ differ from 0, implying that all elements of $\boldsymbol{\beta}_t$ are changing over time. Among the off-diagonal elements, the coefficient of dum_B has meaningful correlation with the coefficient of dum_A and the coefficient of price. Overall, we conclude that in our dataset, $\boldsymbol{\beta}_t$ is apparently time-varying, with a fairly pronounced degree of white noise.

Estimation of $\boldsymbol{\Sigma}_\mathbf{b}$. Table 5 gives estimates for $\boldsymbol{\Sigma}_\mathbf{b}$. Posterior means are again given on and below the main diagonal, with posterior standard deviations in parentheses; the posterior means of the correlation coefficients are given above the main diagonal. This table suggests that $\boldsymbol{\Sigma}_\mathbf{b}$ is neither a null matrix nor a diagonal matrix; furthermore, all diagonal elements are significantly different from 0.

Let us briefly examine the effect of parameter dynamics on the heterogeneity distribution by comparing posterior means for the covariance $\boldsymbol{\Sigma}_\mathbf{b}$ for both the RVAR(1) model and model M_0. For model M_0, the posterior mean for the diagonal of $\boldsymbol{\Sigma}_\mathbf{b}$ is

$$\begin{bmatrix} 5.4701 & 8.4474 & 11.1818 & 1.8029 & 1.9395 & 5.3011 \\ (.6615) & (.9838) & (1.4515) & (.2342) & (.2695) & (.6255) \end{bmatrix},$$

where the respective posterior standard deviations are given in parentheses. These diagonal elements are 18.2–32.0% smaller than the corresponding elements of $\boldsymbol{\Sigma}_\mathbf{b}$ for the RVAR(1) model, which were given in Table 5. Therefore, the traditional random-effects logit model "underestimated" the extent of heterogeneity.

Discussion. The elements of $\mathbf{d}$ and $\mathbf{A}_1$ for the option C dummy are essentially 0, but the corresponding variance in $\boldsymbol{\Sigma}_\mathbf{w}$ is positive. Thus the dynamics for the dummy variable of option C consist of a white noise term only. The dynamics for the dummies for options A and B, on the other hand, constitute AR(1) processes. Allenby and Lenk (1994) also reported an autocorrelated error structure for utilities. Because they introduced a scalar for the error autocorrelation of utilities across choice occasions, they implicitly assumed that choice dummy effects would follow the same type of stochastic process with the sample autocorrelation coefficient. However, our results suggest different stochastic processes for each. Specifically, the feature coefficient seems to follow a pure white noise process, the Display coefficient is found to follow a white noise process with a non-0 mean, and the Price coefficient appears to follow a AR(1) process, over the observation period.

Table 6 gives the posterior means for the following quantities:

- $\boldsymbol{\mu}_\beta$, the long-run mean of $\boldsymbol{\beta}_{ht}$ (and the posterior standard deviation of $\boldsymbol{\mu}_\beta$)
- $\sqrt{\mathrm{var}(\beta_{ht,i})} = \sqrt{\frac{\Sigma_{\mathbf{w},i,i}}{1-A^2_{1,i,i}} + \Sigma_{\mathbf{b},i,i}}$, an overall standard deviation for $\boldsymbol{\beta}_{ht}$
- $\sqrt{\Sigma_{\mathbf{b},i,i}}$, the standard deviation of the heterogeneity component, $\mathbf{b}_h$, of $\boldsymbol{\beta}_{ht}$
- $\sigma_{\beta,i} = \sqrt{\frac{\Sigma_{\mathbf{w},i,i}}{1-A^2_{1,i,i}}}$, the long-run standard deviation of the dynamic component, $\boldsymbol{\beta}_t$, of $\boldsymbol{\beta}_{ht}$
- $\sqrt{\Sigma_{\mathbf{w},i,i}}$, the standard deviation of the "white noise" component of $\boldsymbol{\beta}_t$.

The posterior mean of $\boldsymbol{\mu}_\beta$ displays the anticipated signs. The posterior standard deviations for some of the elements are relatively large, notably for the option C dummy, Feature, and Display, suggesting a fair amount of uncertainty about the actual value of $\boldsymbol{\mu}_\beta$. A comparison of the results for $\boldsymbol{\mu}_\beta$ with those for $\mathbf{d}$ in Table 3 shows a moderate difference for Price.

The overall variability in $\boldsymbol{\beta}_{ht}$, as measured by the posterior mean for the standard deviation $\sqrt{\mathrm{var}(\beta_{ht,i})}$, is quite large. In fact, all of these standard deviations are *larger* than the corresponding elements of $\boldsymbol{\mu}_\beta$, so the corresponding regression coefficients are negative for some households and time periods and positive for others. Thus, although the posterior means for the

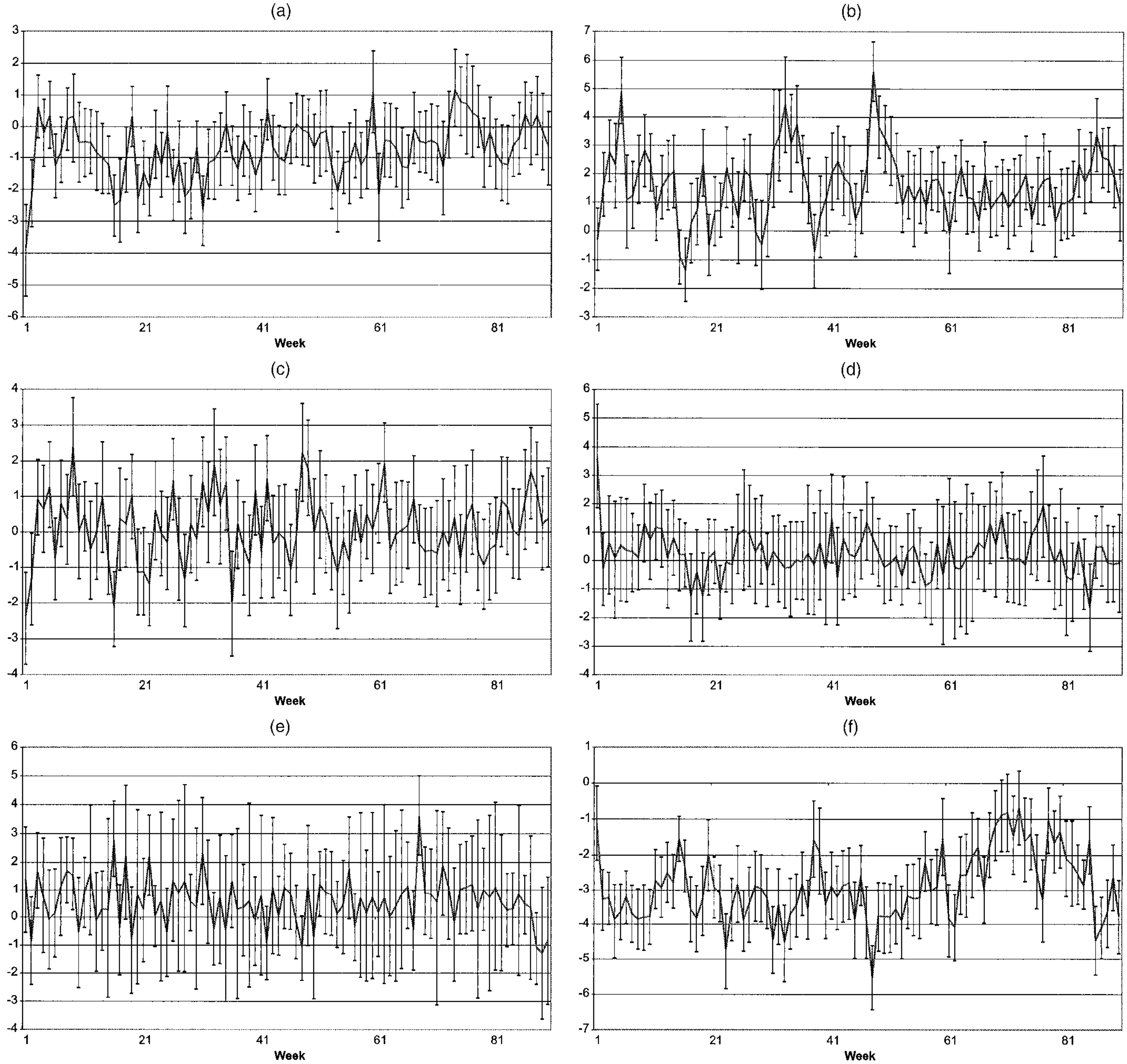

Figure 2. Dynamics of β_t. (a) dum$_A$; (b) dum$_B$; (c) dum$_C$; (d) feature; (e) display; (f) price. Solid lines denote estimated values. The lower and upper bars denote the 5th and 95th percentiles.

elements of the long-run mean μ_β display the anticipated signs, this is not necessarily true for individual households.

Let us next examine the contribution to the overall variability in β_{ht} that can be attributed to household heterogeneity and to parameter dynamics. Household heterogeneity can be measured by the square root of the diagonal elements of $\boldsymbol{\Sigma}_\mathbf{b}$, $\sqrt{\Sigma_{\mathbf{b},i,i}}$, and parameter dynamics can be measured by $\sigma_{\beta,i} = \sqrt{\Sigma_{\mathbf{w},i,i}/(1 - A_{1,i,i}^2)}$; see (21). Table 6 suggests that, except for Feature and Display, the posterior means for the standard deviation of the heterogeneity component are quite a bit larger than the posterior means for the corresponding values of $\sigma_{\beta,i}$. Household heterogeneity is thus a very important component in the overall variability in β_{ht}.

Finally, let us contrast $\sigma_{\beta,i}$, the long-run standard deviation of the dynamic component, $\boldsymbol{\beta}_t = \mathbf{d} + \mathbf{A}_1\boldsymbol{\beta}_{t-1} + \mathbf{w}$, with $\sqrt{\Sigma_{\mathbf{w},i,i}}$, the standard deviation of the "white noise" component, $\mathbf{w}_t$. The value of the posterior mean for $\sqrt{\Sigma_{\mathbf{w},i,i}}$ is only slightly smaller than that for $\sigma_{\beta,i}$. This suggests that the "white noise" component is the dominant force in the parameter dynamics of each component of $\boldsymbol{\beta}_t$.

4.6 Tests for Structural Change

It is important to check whether or not parameter dynamics truly exist in the training sample. After dividing the 90 weeks of data into nine datasets such that $\bar{y}_z = \{y_t\}_{t=10(z-1)+1}^{10z}$, where $z = 1, \ldots, 9$, we estimate all nine regression coefficients $\{\bar{\beta}_z\}_{z=1}^{9}$

Table 3. Estimates of d and A

	Estimate (standard deviation; MC error)	(5th percentile, 95th percentile) interval
d		
dum_A	−.5410 (.2265; .0094)	(−.9182, −.1744)
dum_B	1.1811 (.3020; .0157)	(.6867, 1.6780)
dum_C	.1172 (.2900; .0168)	(−.3634, .5982)
Feature	.1887 (.2213; .0107)	(−.1722, .5536)
Display	.6387 (.3528; .0182)	(.1115, 1.2451)
Price	−2.0741 (.4470; .0243)	(−2.7682, −1.3043)
$diag(A_1)$		
dum_A	.2322 (.1339; .0051)	(.0282, .4632)
dum_B	.2640 (.1397; .0058)	(.0467, .4989)
dum_C	−.0087 (.1525; .0067)	(−.2328, .2655)
Feature	.1659 (.1691; .0134)	(−.1411, .4213)
Display	−.0765 (.1890; .0089)	(−.3999, .2241)
Price	.2958 (.1423; .0068)	(.0747, .5420)

Table 4. Estimate of $\boldsymbol{\Sigma}_{\boldsymbol{w}}$

	dum_A	dum_B	dum_C	Feature	Display	Price
dum_A	2.2014 (.3718)	.1842	.1582	−.0107	−.0581	.0797
dum_B	.4612 (.3033)	2.8483 (.5006)	.3099	−.0106	−.1046	−.2411
dum_C	.3771 (.2864)	.8401 (.3567)	2.5797 (.4453)	.0145	−.0876	−.2210
Feature	−.0249 (.2816)	−.0280 (.3243)	.0364 (.3059)	2.4485 (.4502)	.0129	.0094
Display	−.1580 (.3433)	−.3238 (.3942)	−.2581 (.3784)	.0371 (.3661)	3.3642 (.6998)	.0715
Price	.1756 (.2596)	−.6038 (.3089)	−.5267 (.2931)	.0218 (.2783)	.2000 (.3354)	2.2031 (.3630)

Table 5. Estimate of $\boldsymbol{\Sigma}_{\boldsymbol{b}}$

	dum_A	dum_B	dum_C	Feature	Display	Price
dum_A	7.2242 (.8232)	.1683	.2161	−.1588	−.0699	−.0151
dum_B	1.5397 (.7671)	11.5826 (1.3067)	.2263	−.0486	−.1414	−.4943
dum_C	2.3549 (.9208)	3.1222 (1.1404)	16.4344 (2.0409)	−.0314	−.1346	−.2889
Feature	−.6407 (.4182)	−.2483 (.5370)	−.1913 (.6885)	2.2548 (.3371)	−.0736	.0311
Display	−.2898 (.4220)	−.7423 (.5843)	−.8419 (.7388)	−.1706 (.2292)	2.3795 (.3906)	.1282
Price	−.1034 (.5319)	−4.2839 (.7081)	−2.9822 (.7751)	.1189 (.3935)	.5036 (.4069)	6.4841 (.7261)

Table 6. Elements of Variation in β_{ht}

i	μ_β	$\sqrt{var(\beta_{ht,i})}$	$\sqrt{\Sigma_{\boldsymbol{b},i,i}}$	$\sigma_{\beta,i}$	$\sqrt{\Sigma_{\boldsymbol{w},i,i}}$
dum_A	−.7037 (.2739)	3.0965 (.1523)	2.6834 (.1528)	1.5385 (.1433)	1.4786 (.1235)
dum_B	1.6095 (.3129)	3.8334 (.1910)	3.3979 (.1920)	1.7667 (.1664)	1.6813 (.1462)
dum_C	.1147 (.2882)	4.3613 (.2407)	4.0462 (.2499)	1.6201 (.1415)	1.6003 (.1365)
Feature	.2270 (.2710)	2.2001 (.1423)	1.4975 (.1115)	1.6067 (.1551)	1.5583 (.1416)
Display	.5849 (.2909)	2.4236 (.1809)	1.5375 (.1247)	1.8671 (.2023)	1.8246 (.1873)
Price	−2.9491 (.2833)	2.9930 (.1504)	2.5424 (.1423)	1.5728 (.1518)	1.4794 (.1206)

for these nine datasets simultaneously, where $\bar{\beta}_z$ contains logit coefficients belonging to $\bar{y}_z$. As before, to estimate $\{\bar{\beta}_z\}_{z=1}^{9}$, we use the MCMC sampler; specifically, distributions involving β in (14) are changed to estimate $\{\bar{\beta}_z\}_{z=1}^{9}$ as follows:

$$\left(\prod_{z=1}^{9} p(\bar{y}_z|\bar{\beta}_z, \mathbf{b})\right) p(\bar{\beta}_z). \tag{23}$$

It is readily apparent that the foregoing model is a counterpart to the tests on structural intercept and slope changes in the classical econometrics literature. Hence we call (23) a *structural change* model. By estimating this model, we can test

$$\mathrm{H}_1: \bar{\beta}_1 = \cdots = \bar{\beta}_9 \quad (M_0)$$

and

$$\mathrm{H}_2: \bar{\beta}_1 \neq \cdots \neq \bar{\beta}_9 \quad \text{(structural change model)}.$$

The log of the integrated likelihood of the structural change model is $-3{,}351.46$. Clearly, the null hypothesis H_1 is rejected (Bayes factor favoring H_1 over $\mathrm{H}_2 = 1.38e-37$). This in turn further verifies that parameter dynamics exist for these data. Note that the RVAR(1) model is still decisively preferred over the structural change model [Bayes factor favoring RVAR(1) over the structural change model $= 5.59e+135$], implying that β_t varies within each set of observations.

4.7 Aggregation Bias

The estimation results of the structural change model raise an issue. Typically, a researcher uses a subset of the entire available data for model estimation purposes. However, the assumption that information obtained from currently available data will also be valid in the future may be problematic; furthermore, obtained estimates can depend on the time periods for which a choice model is fitted. Thus if parameter dynamics exist, then estimates deriving from M_0 can suffer from aggregation bias.

To illustrate the potential for aggregation bias, we compare the estimates of the structural change model with those of both M_0 and RVAR(1). As shown in Figure 3, there are several cases in which the estimates of M_0 deviate noticeably from the estimates of the structural change model. However, the estimates of RVAR(1) strongly overlap with those of the structural change model. This suggests that it may be possible to substantially reduce the degree of aggregation bias if parameter dynamics are appropriately accounted for.

4.8 Effects of Parameter Dynamics on Choice Behavior

We have shown that by incorporating temporal variation in parameters directly, choice dynamics can be better captured through a form of VAR process than by either the traditional static model or previous dynamic models. To examine potential sources of superior prediction of choice dynamics, we now investigate the effects of exogenous covariates and parameter dynamics on choice behavior.

From (2), define the following derivatives:

$$\rho^h_{x_{t,k}}(j,i) = \frac{\partial p_{hjt}}{\partial x_{hit,k}}, \qquad \rho^h_{\beta_{t,k}}(j) = \frac{\partial p_{hjt}}{\partial \beta_{t,k}},$$

and

$$\rho^h_{x_{t,k}\beta_{t,k}}(j,i) = \frac{\partial^2 p_{hjt}}{\partial x_{hit,k}\,\partial \beta_{t,k}}.$$

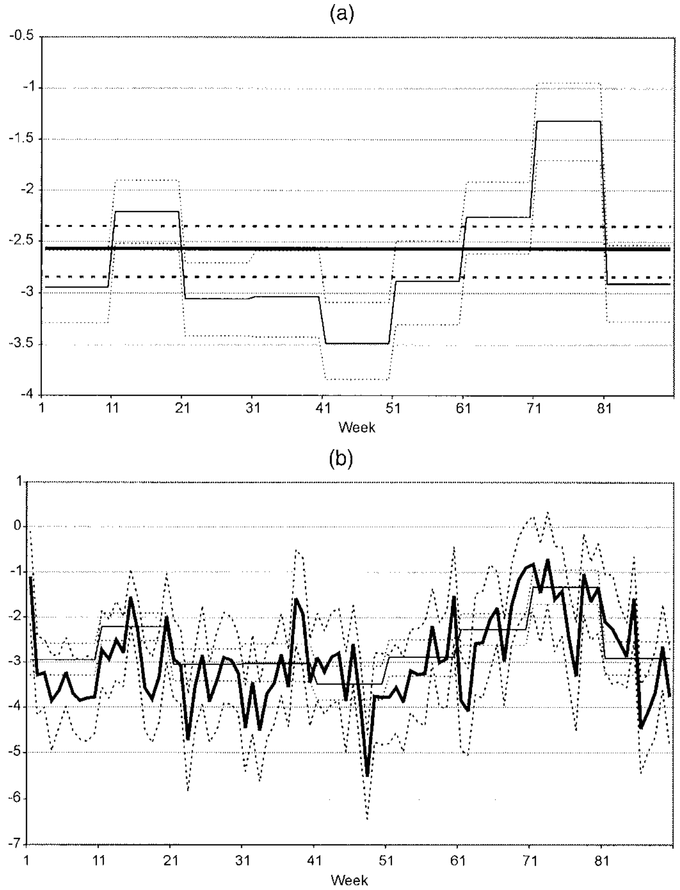

Figure 3. Comparison of Price Coefficient Estimates for the Structural Change (—) and RVAR(1) (—) Models. (a) Structural change versus M_0; (b) structural change versus RVAR(1). Lower and upper dotted lines enveloping each solid line of estimates denote the corresponding 5th and 95th percentiles.

Next, we computed $\rho_{x_{t,k}}(j,i) = \frac{1}{n_t}\sum_{h\in\mathcal{H}_t}\rho^h_{x_{t,k}}(j,i)$ as the sample average of $\rho^h_{x_{t,k}}(j,i)$ in period t, where n_t is the sample size of $\mathcal{H}_t$. Similarly, we computed the following sample averages of the foregoing quantities: $\rho_{x_{t,k}}(j,j)$, $\rho_{\beta_{t,k}}(j)$, $\rho_{x_{t,k}\beta_{t,k}}(j,j)$, and $\rho_{x_{t,k}\beta_{t,k}}(j,i)$. For M_0, $\beta_{t,k}$ is replaced by the kth element of the regression coefficients. We compute these sample average estimates for both M_0 and RVAR(1) for each time period. For option $j=A$ and variable $=$ Price ($k=6$), Table 7 gives the MCMC estimates of these household-averaged derivatives further averaged over the 90-week observation period, for example, $\bar{\rho}_{x_{t,k}}(j,i) = \frac{1}{90}\sum_{t=1}^{90}\rho_{x_{t,k}}(j,i)$. The pattern of results indicates that M_0 tends to overestimate all quantities of interest.

4.9 m-Step-Ahead Parameter Forecasting

Let us now consider forecasting. Given that we have data through period T, m-step-ahead forecasts can be readily obtained with another MCMC run. For example, such a simulation yields posterior distributions for $\boldsymbol{\beta}_{T+m}$.

We conducted six-step-ahead forecasting, obtaining posterior distributions for $\boldsymbol{\beta}_{T+m}$, where $T=90$ and $m=1,\ldots,6$. At each MCMC iteration, it is straightforward to simulate these future parameters, $\boldsymbol{\beta}_{T+m}$, given $\boldsymbol{\beta}$, $\mathbf{d}$, $\mathbf{A}_1$, and $\boldsymbol{\Sigma}_{\mathbf{w}}$, using (5). The posterior means and standard deviations of $\boldsymbol{\beta}_{T+m}$ are, therefore, readily available from MCMC runs of $\boldsymbol{\beta}_{T+m}$. Table 8 gives

Table 7. The Averaged Effects of Covariate and Parameters on Choice Behavior

	M_0	RVAR(1)
$\bar{\rho}_{x_{t,k}}(A,A)$	−.1515 (.0047)	−.1408 (.0052)
$\bar{\rho}_{x_{t,k}}(A,B)$	.0616 (.0028)	.0556 (.0030)
$\bar{\rho}_{x_{t,k}}(A,C)$	.0333 (.0021)	.0324 (.0023)
$\bar{\rho}_{x_{t,k}}(A,D)$	.0565 (.0027)	.0528 (.0028)
$\bar{\rho}_{\beta_{t,k}}(A)$	−.0461 (.0015)	−.0383 (.0013)
$\bar{\rho}_{\beta_{t,k}}(B)$	.0404 (.0015)	.0327 (.0013)
$\bar{\rho}_{\beta_{t,k}}(C)$	.0227 (.0012)	.0191 (.0011)
$\bar{\rho}_{\beta_{t,k}}(D)$	−.0170 (.0013)	−.0135 (.0011)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}(A,A)$	.0559 (.0023)	.0447 (.0022)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}(A,B)$	−.0151 (.0014)	−.0119 (.0013)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}(A,C)$	−.0110 (.0012)	−.0087 (.0011)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}(A,D)$	−.0298 (.0011)	−.0241 (.0011)

NOTE: Standard deviations are in parentheses.

their means and standard deviations for weeks 91–96, suggesting considerable forecasting uncertainty.

We compared the performance of the six forecasted $\boldsymbol{\beta}_{T+m}$ values based on the RVAR(1) model with that of the traditional static logit model M_0. For both models, we computed log-likelihood values for the prediction dataset. We obtained these likelihood values by first computing predicted choice probabilities, (2), for the forecasting dataset given all parameters simulated at each MCMC iteration, then taking the average of these predicted choice probabilities across MCMC iterations. The computed log-likelihoods for the forecasting sample, the summation of log of (2) given the average of choice probabilities, are −202.21 for RVAR(1) and −208.89 for M_0. For the future parameter forecasting sample, RVAR(1) demonstrates slightly better performance than M_0.

5. CONCLUSION AND FUTURE RESEARCH

Although choice models have achieved a great deal of sophistication over the past decade, researchers have only recently begun to address the interplay of choice dynamics and parameter dynamics. To this end, we have proposed a general VAR framework to account for the phenomenon, one that can be grafted onto any specifications for utility or error structure. In this framework, we can rigorously test a number of hypotheses about the nature of parametric evolution—among them its order, which parameters are involved, and which affect others—as well as demonstrate improved predictive performance.

A number of clear conclusions emerge from our empirical analysis. First and foremost, some (although not all) of the parameters demonstrated strong evidence of temporal variation. This was clear even under the parsimonious specification that emerged as the strongest candidate, RVAR(1). Incorporating such a stochastic parametric structure into existing models would entail a comparatively modest increase in the number of estimated quantities, and should emerge as an attractive alternative to models presuming parametric constancy. Second, forecast performance was improved substantially over the standard random-effects logit model. In fact, the random-effects model appears prone to aggregation biases when its parameter estimates deviate from the implied long-term levels suggested by the VAR(p) specification. To our knowledge, this result is new, and we believe that it merits study in and of itself, given the popularity of the random-effects logit modeling framework. Finally, our data suggest that choice dynamics may be misattributed to exogenous covariates when parameters are presumed not to have dynamics of their own. For example, the random-effects model appears to underadjust for brand-switching behavior, perhaps because such behavior is assumed to be governed by external stimuli, given fixed parameters.

With respect to possible explanations for parameter dynamics, a number of potential explanations can be ruled out—specifically, systematic changes in the characteristics of pooled samples over time and changes in the distribution of stimuli across options. Further, analytic examination and simulation demonstrated that if all or some of the population update their parameters over time, then systematic parameter dynamics may exist even at the population level, as captured by the VAR(p) process.

Suggesting explanations for parametric evolution post hoc, other than those already tested, amounts to speculation. Some authors, however, have provided bases for further investigations along these lines. Yang, Allenby, and Fennell (2002) noted that scanner panel data do not accommodate the proper unit of analysis in modeling preference changes: a person-activity occasion. They discussed how for many activities (e.g., snacking, serving wine), the consumer environment is not constant from one usage occasion to another, so that preferences are rightly situationally or motivationally dependent. Although they explicitly pointed out that occasions for use of laundry detergent,

Table 8. Six-Step-Ahead Forecasting of $\boldsymbol{\beta}_t$

	dum_A	dum_B	dum_C	Feature	Display	Price
Week 91	−.6915	1.4296	.1153	.1756	.6704	−3.1938
	(.2841)	(.3402)	(.3081)	(.3335)	(.4910)	(.3367)
Week 92	−.7000	1.5482	.1206	.2119	.5336	−3.0395
	(.2620)	(.3015)	(.2883)	(.2686)	(.3057)	(.2871)
Week 93	−.7024	1.5858	.1151	.2223	.5946	−2.9868
	(.2671)	(.3037)	(.2880)	(.2672)	(.3020)	(.2797)
Week 94	−.7032	1.5993	.1151	.2252	.5790	−2.9664
	(.2705)	(.3074)	(.2881)	(.2686)	(.2895)	(.2798)
Week 95	−.7035	1.6048	.1147	.2263	.5868	−2.9576
	(.2721)	(.3097)	(.2881)	(.2698)	(.2930)	(.2808)
Week 96	−.7036	1.6072	.1147	.2267	.5837	−2.9535
	(.2729)	(.3110)	(.2881)	(.2703)	(.2901)	(.2817)
Long-term mean	−.7037	1.6095	.1147	.2270	.5849	−2.9491

NOTE: Standard deviations are in parentheses.

the product class used in our study, are less likely to be subject to this sort of temporal preference variation, we believe that their approach merits formal study on data like our own, which would provide the proverbial strong test. One would need recourse to purchase occasion data transcending the panel record alone, and Yang et al. presented approaches to this practical problem at length.

In a similar vein, Wakefield and Inman (2003) also noted that little research has focused on the effects of consumption occasion or context on consumer price sensitivity. They found price sensitivity to be attenuated by hedonic and social consumption situations; because intended consumption occasion varies across consumers and time, this variation is unobserved and could well lead to a moderate degree of parametric evolution in some categories. There is also the related issue of seasonality, although product usage cycles for most frequently purchased goods are considerably shorter than can be supported by purely seasonal explanation. We believe that such issues can be addressed directly through access to auxiliary data—surveys, logs, or self-reports—on individual panelist's usage occasions, perhaps supplemented by brand-by-brand household-level stocks. Such data allow for a modeling framework that accounts for parametric evolution at a less-aggregated, perhaps individual, level. Implementing such a model presents substantial challenges in terms of both data requirements and estimation technology, although we suspect each of these impediments to wane with time.

Our model is not without its limitations. One such limitation is the requirement for data over a relatively long period. In many applications, particularly in field data, long strings of choices are not often available. Another limitation involves variable selection. To be sure, this problem bedevils all empirical choice research, but we know little about the dependence of the present model, in terms of order selection for p, on the choice of covariates. Finally, the model itself can entail a very large number of parameters, making model comparison and interpretation considerably more challenging.

Limitations aside, the model can be widely applied in choice research, due to both its generality and its silence on utility and error structure. We believe that it can be readily extended to include parameter dynamics on an individual level or in a mixture modeling framework. Such an extension would allow different groups of decision makers to update their sensitivities in different ways and would, in our view, offer another compelling dimension through which to examine varied choice behavior.

[Received February 2002. Revised April 2004.]

REFERENCES

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

——— (1995), "Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice," *Journal of Business & Economic Statistics*, 13, 281–289.

Barnett, G., Kohn, R., and Sheather, S. (1996), "Bayesian Estimation of an Autoregressive Model Using Markov Chain Monte Carlo," *Journal of Econometrics*, 74, 237–254.

Bernardo, J. M., and Smith, A. F. M. (1994), *Bayesian Theory*, New York: Wiley.

Cargnoni, C., Müller, P., and West, M. (1997), "Bayesian Forecasting of Multinomial Time Series Through Conditionally Gaussian Dynamic Models," *Journal of the American Statistical Association*, 92, 640–647.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493–500.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335.

Doan, T., Litterman, R. B., and Sims, C. A. (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–144.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to Calculation of Posterior Moments," in *Bayesian Statistics 4*, eds. J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 169–193.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society*, Ser. B, 38, 205–247.

Heckman, J. J. (1981), "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data With Econometric Applications*, eds. C. F. Manski and D. McFadden, Cambridge, MA: MIT Press, pp. 179–195.

Heidelberger, P., and Welch, P. D. (1983), "Simulation Run Length Control in the Presence of an Initial Transient," *Operations Research*, 31, 1109–1144.

Jenkins, G. M., and Watts, D. G. (1968), *Spectral Analysis and Its Applications*, San Francisco: Holden-Day.

Kahneman, D., and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–291.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Li, H., and Tsay, R. S. (1998), "A Unified Approach to Identifying Multivariate Time Series Models," *Journal of the American Statistical Association*, 93, 770–782.

Litterman, R. B. (1986), "Forecasting With Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business & Economic Statistics*," 4, 25–38.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Heidelberg: Springer-Verlag.

Manski, C. F. (1977), "The Structure of Random Utility Models," *Theory and Decision*, 8, 229–254.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. P. Zarembka, New York: Academic Press, pp. 105–142.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of State Calculations by Fast Computing Machine," *Journal of Chemical Physics*, 21, 1087–1091.

Neal, R. M. (1997), "Markov Chain Monte Carlo Methods Based on 'Slicing' the Density Function," Technical Report 9722, University of Toronto, Dept. of Statistics.

Newton, M. A., and Raftery, A. E. (1994), "Approximate Bayesian Inference With the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society*, Ser. B, 56, 3–48.

Polasek, W., and Kozumi, H. (1996), "The VAR–VARCH Model: A Bayesian Approach," in *Modelling and Prediction: Honoring Seymour Geisser*, eds. J. C. Lee, W. O. Johnson, and A. Zellner, New York: Springer-Verlag, pp. 402–413.

Roberts, G. O., and Rosenthal, J. S. (1999), "Convergence of Slice Sampler Markov Chains," *Journal of the Royal Statistical Society*, Ser. B, 61, 643–660.

Slovic, P., Griffin, D., and Tversky, A. (1990), "Compatibility Effects in Judgement and Choice," in *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, ed. R. M. Hogarth, Chicago: University of Chicago Press, pp. 5–27.

Thaler, R. (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4, 199–214.

Tversky, A., Sattath, S., and Slovic, P. (1988), "Contingent Weighting in Judgement and Choice," *Psychological Review*, 95, 371–384.

von Nitzsch, R., and Weber, M. (1993), "The Effect of Attribute Ranges on Weights in Multiattribute Utility Measurements," *Management Science*, 39, 937–943.

Wakefield, K. L., and Inman, J. J. (2003), "Situational Price Sensitivity: The Role of Consumption Occasion, Social Context, and Income," *Journal of Retailing*, 79, 199–212.

West, M., and Harrison, P. (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag.

Yang, S., Allenby, G. M., and Fennell, G. (2002) "Modeling Variation in Brand Preference: The Roles of Objective Environment and Motivating Conditions," *Marketing Science*, 21, 14–31.