Modeling Parametric Evolution in a Random Utility Framework

Jin Gyo Kim, Ulrich Menzefricke and Fred M. Feinberg

21 April, 2004

Jin Gyo Kim (kimjg@mit.edu), MIT Sloan School of Management, 38 Memorial Drive, E56-323, Cambridge, MA, 02142; Ulrich Menzefricke (menzefricke@rotman.utoronto.ca), Rotman School of Management, University of Toronto, 105 St. George St., Toronto, Ontario, Canada, M5S 3E6; and Fred Feinberg (feinf@umich.edu), University of Michigan Business School, 701 Tappan St., Ann Arbor, Michigan, USA, 48109.

Forthcoming at *Journal of Business and Economic Statistics*.

Modeling Parametric Evolution in a Random Utility Framework

Abstract

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 additionally modeling temporal changes in parameters directly, 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.

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 to. A rich literature has emerged to aid researchers in linking exogenous covariates to temporal changes in choices. Because data are available on observed choices, and not on unobserved measures of relative 'attractiveness' of available options, the dominant method of achieving such a linkage has been that of Random Utility models (McFadden 1973; Manski 1977).

Within the random utility framework, one need 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 modeling changes in parameters directly. 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 we develop will indeed distinguish parameters along these lines, and use that knowledge for improved forecasting. The paper is organized as follows. We review prior 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 attributes depends on such contextual features as scale compatibility (Slovic *et al.* 1990), strategy compatibility (Tversky *et al.* 1988), and framing (Kahneman and Tversky 1981; 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 capture 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. By 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 $\alpha^{'}$ 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 will use three generic subscripts: h denotes an individual unit of observation ($h = 1, ..., H$), j denotes an option ($j = 1, ..., J$), and t denotes time ($t = 1, ..., T$). Let $y_{ht} = j$ denote the event that unit h chooses option j at time t, x_{hjt} denote unit h's k-dimensional covariate vector for unit j at time t, and u_{hjt} denote unit h's utility for option j at time t. Thus

$$u_{hjt} = \beta_{ht}^{'} x_{hjt} + \varepsilon_{hjt}, \tag{1}$$

where where β_{ht} is the k-dimensional coefficient vector for unit h at time t and ε_{hjt} the associated error. If ε_{hjt} are iid Gumbel, then a dynamic logit model arises for choice probability p_{hjt} (cf.

McFadden 1973):

$$p_{hjt} = p(y_{ht} = j|\beta_{ht}) = \frac{\exp({\beta_{ht}}' x_{hjt})}{\sum_{i=1}^{J} \exp({\beta_{ht}}' x_{hit})}. \tag{2}$$

To model parametric temporal variation, we assume β_{ht} can be decomposed into two parts,

$$\beta_{ht} = \beta_t + b_h, \tag{3}$$

where β_t is a time-varying coefficient vector common across units, and b_h is a vector of random effects to incorporate heterogeneity *across* units. (2) now simplifies to

$$p_{hjt} = p(y_{ht} = j|\beta_t, b_h). \tag{4}$$

To capture dynamics for β_t, we introduce a vector autoregressive process of order p, VAR(p) (Li and Tsay 1998; Lütkepohl 1991; Polasek and Kozumi 1996),

$$\begin{aligned} \beta_t &= d + \sum_{n=1}^{p} A_n \beta_{t-n} + w_t,\ w_t \sim N(0, \Sigma_w),\ t = 1, \ldots, T \\ &= d + AZ_{t-1} + w_t, \end{aligned} \tag{5}$$

where d is a k-dimensional vector, A_n is a $(k \times k)$ coefficient matrix, $A = (A_1, ..., A_p)$ is a $k \times kp$ matrix, $Z_t = (\beta_t^{'}, ..., \beta_{t-p+1}^{'})^{'}$ is a kp-dimensional vector, and 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). Since A_n is not assumed 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 β_t and a different element of β_{t-n}. Further, (5) can capture stationary as well as non-stationary dynamics; it is well-known that the VAR(p) process is stable and thus stationary if

$$\det(I_k - A_1 l - ... - A_p l^p) \neq 0 \text{ for } |\ l\ | \leq 1, \tag{6}$$

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, the expected value of β_t does not depend on t, that is,

$$\mu_\beta = E(\beta_t) = (I - A_1 - A_2 - ... - A_p)^{-1} d,$$

where the expectation is with respect to w_t, $t = 1, 2, \ldots$ (Lütkepohl 1991).

Finally, we model the heterogeneity of b_h in (3) as a multivariate normal random effect,

$$p(b_h|\Sigma_b) = N_k(0, \Sigma_b), \ \forall\ h, \tag{7}$$

where Σ_b is an unknown covariance matrix.

2.3 Prior Distributions

Priors are required for $\{\beta_t\}_{t=1-p}^{0}$, d, A, Σ_w and Σ_b. Following standard assumptions of dynamic state-space models (Cargnoni *et al.* 1997; Carlin *et al.* 1992; Harrison and Stevens 1976; West and Harrison 1997), we assume that $\{\beta_t\}_{t=1-p}^{0}$, d, A, Σ_w, and Σ_b are mutually independent, and use the following prior distributions,

$$p(\beta_i) = N_k(m_0, S_0), \text{ where } i = 1-p, ..., 0, \tag{8}$$

$$p(d) = N_k(m_d, S_d), \tag{9}$$

$$p(vec(A)) = N_{k^2p}(m_\alpha, S_\alpha), \tag{10}$$

$$p(\Sigma_w) = IW_k(v_w, S_w), \text{ and} \tag{11}$$

$$p(\Sigma_b) = IW_k(v_b, S_b). \tag{12}$$

Here $vec(\cdot)$ is the usual column stacking operator so that $vec(A)$ is a k^2p-dimensional vector. The expression $p(\Sigma) = IW_k(v, S)$ denotes that Σ has a k-dimensional inverted Wishart distribution with parameters v and S, where $v > 0$ and S is non-singular, that is, $p(\Sigma) = IW_k(v, S) \propto |\Sigma|^{-(\frac{1}{2}v+k)} \exp(-\frac{1}{2}\text{tr } \Sigma^{-1}S)$. Furthermore, the parameters of the prior distributions (m_0, S_0, m_d, S_d, m_α, S_α, v_w, S_w, v_b, S_b) are known values, which we choose to obtain non-informative proper priors. In particular, for the prior of $vec(A)$, we use the Minnesota priors (Doan *et al.* 1984; Litterman 1986), which are specialized for the VAR(p) process. Specifically, we choose $m_\alpha = 0$ and S_α to be a diagonal matrix with elements:

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

where $s_{rc,n}$ is the prior variance of the (r, c) element of A_n $(n = 1, \ldots, p)$, σ_r/σ_c is the ratio of square roots of corresponding diagonal elements of Σ_w, λ is the prior belief on the tightness around zero for the diagonal elements of A_1, and $0 < \theta < 1$ reflects prior belief that most of the variation of β_t is explained by its own lags. Thus, the Minnesota priors are locally non-informative proper priors around zero, an attractive property because, under the stability condition, 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 and 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

- $\mathcal{H} = \{1, 2, ..., H\}$ 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,
- $y_t = \{y_{ht}\}_{h \in \mathcal{H}_t}$ denote the observed choice data at time t,
- $y = (y_1^{'}, ..., y_T^{'})^{'}$ denote all choice data from time 1 to time T,
- $\beta = (\beta_1^{'}, ..., \beta_T^{'})^{'}$, $b_t = \{b_h\}_{h \in \mathcal{H}_t}$, and $b = \{b_h\}_{h \in \mathcal{H}}$.

Then, the posterior distribution is

$$\begin{aligned} p(\beta, b, d, A, \Sigma_w, \Sigma_b | y) \quad \propto \quad & \left(\prod_{t=1}^{T} p(y_t | \beta_t, b_t) \right) \times \left(\prod_{n=1-p}^{0} p(\beta_n) \right) \times \\ & \left(\prod_{t=1}^{T} p(\beta_t | d, A, \beta_{t-1}, ..., \beta_{t-p}, \Sigma_w) \right) \times \\ & \left(\prod_{\mathcal{H}} p(b_h | \Sigma_b) \right) \times p(d) \times p(A) \times p(\Sigma_w) \times p(\Sigma_b), \end{aligned} \tag{14}$$

where,

$$p(y_t | \beta_t, b_t) = \prod_{h \in \mathcal{H}_t} \prod_{j=1}^{J} p_{hjt}^{q_{hjt}},$$

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 β_t $(t = 1 - p, ..., T)$ there are $k(p + T)$; for b_h $(h = 1, ..., H)$, kH; for d, k; for A, $k^2 p$; and for both Σ_w and Σ_b, $\frac{1}{2}k(k+1)$. In the forthcoming illustration, we will have $H = 492$, $T = 90$ and $k = 6$, yielding $3540 + 42p$ elements overall.

Because analytic methods are not available to evaluate the posterior distribution in (14), we employ Markov chain Monte Carlo (MCMC) methods, as described in Sec. 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:

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

where $p(\bullet|H_i)$ is the prior on parameters under model i ($i = 1, 2$) and $p(y|\bullet, H_i)$ 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 $\Phi^{(g)}, g = 1, ..., G$, denote the G values of Φ generated from the posterior distribution of Φ, $p(\Phi|H_1)$. The integrated likelihood for Model 1, $p(y|H_1) = \int p(y|\Phi, H_1)p(\Phi|H_1)d\Phi$ in (15), can be estimated by the harmonic mean estimator (Newton and Raftery 1994),

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

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 MCMC Sampler

To evaluate the posterior distribution, $p(\beta, b, d, A, \Sigma_w, \Sigma_b|y)$, given in (14), we implement a MCMC sampler, using the following conditional posterior distributions,

$p(\beta|b, d, A, \Sigma_w, \Sigma_b, y) \leftrightarrow p(b|\beta, d, A, \Sigma_w, \Sigma_b, y) \leftrightarrow p(d|\beta, A, \Sigma_w, \Sigma_b, y) \leftrightarrow$
$p(A|\beta, b, d, \Sigma_w, \Sigma_b, y) \leftrightarrow p(\Sigma_w|\beta, b, d, A, \Sigma_b, y) \leftrightarrow p(\Sigma_b|\beta, b, d, A, \Sigma_w, y)$.

We next describe the sampling procedure for each.

3.2.1 Sampling from $p(\beta|b, d, A, \Sigma_w, \Sigma_b, y)$

To sample from $p(\beta|b, d, A, \Sigma_w, \Sigma_b, y)$, we need the conditional posterior density for β_t, $p(\beta_t|\beta_{m\neq t}, b, d, A, \Sigma_w, \Sigma_b, y)$. When $1 \leq t \leq T$,

$$p(\beta_t|\beta_{m\neq t}, b, d, A, \Sigma_w, \Sigma_b, y) \quad \propto \quad \prod_{m=0}^{p} p(\beta_{t+m}|d, A, \beta_{t+m-1}, ..., \beta_{t+m-p}, \Sigma_w)$$

$$\times p(y_t|\beta_t, b_t) \tag{16}$$

$$= \quad N(F_t f_t, F_t) p(y_t|\beta_t, b_t), \tag{17}$$

where

$$F_t^{-1} = \begin{cases} S_0^{-1} + \sum_{m=1-t}^{p} A_m^{'} \Sigma_w^{-1} A_m, \ t = 1-p, ..., -1, 0, \\ \Sigma_w^{-1} + \sum_{m=1}^{p} A_m^{'} \Sigma_w^{-1} A_m, \ t = 1, ..., T-p, \\ \Sigma_w^{-1} + \sum_{m=1}^{T-t} A_m^{'} \Sigma_w^{-1} A_m, \ t = T-p+1, ..., T-1, \\ \Sigma_w^{-1}, \ t = T, \end{cases}$$

and

$$f_t^{'} = \begin{cases} m_0^{'} S_0^{-1} + \sum_{m=1-t}^{p} (\beta_{t+m} - d - \sum_{n=1, n \neq m}^{p} A_n \beta_{t+m-n})^{'} \Sigma_w^{-1} A_m^{'}, \ t = 1-p, ..., -1, 0, \\ (d + \sum_{m=1}^{p} A_m \beta_{t-m})^{'} \Sigma_w^{-1} + \sum_{m=1}^{p} (\beta_{t+m} - d - \sum_{n=1, n \neq m}^{p} A_n \beta_{t+m-n})^{'} \Sigma_w^{-1} A_m^{'}, \ t = 1, ..., T-p, \\ (d + \sum_{m=1}^{p} A_m \beta_{t-m})^{'} \Sigma_w^{-1} + \sum_{m=1}^{T-t} (\beta_{t+m} - d - \sum_{n=1, n \neq m}^{p} A_n \beta_{t+m-n})^{'} \Sigma_w^{-1} A_m^{'}, \ t = T-p+1, ..., T-1, \\ (d + \sum_{m=1}^{p} A_m \beta_{T-m})^{'} \Sigma_w^{-1}, \ t = T. \end{cases}$$

Given (17), a Metropolis-Hastings algorithm step can be conducted as follows (Chib and Greenberg 1995; Metropolis *et al.* 1953):

1. Sample β_t^* from a proposal density, $N(\beta_t^{pre}, \phi_\beta I)$, where β_t^{pre} is the most recently updated value and ϕ_β is a fixed tuning constant.

2. Substitute β_t^* for β_t^{pre} with acceptance probability

$$\pi(\beta_t^*, \beta_t^{pre}) = \min \left(\frac{p(y_t|\beta_t^*, b_t) n(\beta_t^*|F_t f_t, F_t)}{p(y_t|\beta_t^{pre}, b_t) n(\beta_t^{pre}|F_t f_t, F_t)}, 1 \right),$$

   where $N(\beta_t|F_t f_t, F_t)$ denotes the multivariate normal density with mean $F_t f_t$ and covariance matrix F_t, evaluated at β_t.

3.2.2 Sampling from $p(b|\beta, d, A, \Sigma_w, \Sigma_b, y)$ and $p(d|\beta, A, \Sigma_w, \Sigma_b, y)$

The conditional posterior density for b_h is

$$p(b_h|\beta_t, d, A, \Sigma_w, \Sigma_b, y_{ht}) \propto p(b_h|\Sigma_b) \prod_{t_h} p(y_{ht}|\beta_t, b_h),$$

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

1. Sample b_h^* from a proposal density, $N(b_h^{pre}, \phi_b I_k)$, where b_h^{pre} is the most recently updated value and ϕ_b is a fixed tuning constant.

2. Substitute b_h^* for b_h^{pre} with acceptance probability

$$\pi(b_h^*, b_h^{pre}) = \min \left(\frac{p(b_h^*|\Sigma_b) \prod_{t_h} p(y_{ht}|\beta_t, b_h^*)}{p(b_h^{pre}|\Sigma_b) \prod_{t_h} p(y_{ht}|\beta_t, b_h^{pre})}, 1 \right).$$

The conditional posterior density for d is

$$
\begin{aligned}
p(d|\beta, A, \Sigma_w, \Sigma_b, y) &\propto \prod_{t=1}^{T} p(\beta_t|d, A, Z_{t-1}, \Sigma_w)p(d) \\
&= N(\mu_d^*, \Sigma_d^*),
\end{aligned} \tag{18}
$$

a multivariate normal density with mean vector $\mu_d^* = \Sigma_d^*\{S_d^{-1}m_d + \sum_{t=1}^{T}\Sigma_w^{-1}(\beta_t - AZ_{t-1})\}$ and covariance matrix $\Sigma_d^* = (S_d^{-1} + T\Sigma_w^{-1})^{-1}$.

3.2.3 Sampling from $p(A|\beta, b, d, \Sigma_w, \Sigma_b, y)$

Let the (k^2p)-vector $\boldsymbol{\alpha} = vec(A)$, let $Z = (Z_0, Z_1, ..., Z_{T-1})$, $\tilde{\beta}_t = \beta_t - d$, $\tilde{\beta} = (\tilde{\beta}_1, ..., \tilde{\beta}_T)$, and let the kT-vector $\tilde{\boldsymbol{\beta}} = vec(\tilde{\beta})$. Then, the conditional posterior of $\boldsymbol{\alpha}$ is

$$
\begin{aligned}
p(\boldsymbol{\alpha}|\tilde{\boldsymbol{\beta}}, \Sigma_w) &= p(\boldsymbol{\alpha}|\beta, b, d, \Sigma_w, \Sigma_b, y) \propto p(\tilde{\boldsymbol{\beta}}|\boldsymbol{\alpha}, \Sigma_w)p(\boldsymbol{\alpha}) \\
&\propto \exp[-\frac{1}{2}\{(\tilde{\boldsymbol{\beta}} - (Z^{'} \otimes I_k)\boldsymbol{\alpha})^{'}(I_T \otimes \Sigma_w^{-1})(\tilde{\boldsymbol{\beta}} - (Z^{'} \otimes I_k)\boldsymbol{\alpha}) \\
&\quad +(\boldsymbol{\alpha} - m_\alpha)^{'}S_\alpha^{-1}(\boldsymbol{\alpha} - m_\alpha)\}].
\end{aligned}
$$

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

$$
p(\boldsymbol{\alpha}|\tilde{\boldsymbol{\beta}}, \Sigma_w) = N(\boldsymbol{\alpha}^*, \Sigma_{\boldsymbol{\alpha}}^*), \tag{19}
$$

a (k^2p)-dimension normal density with mean vector $\boldsymbol{\alpha}^* = \Sigma_{\boldsymbol{\alpha}}^*\{S_\alpha^{-1}m_\alpha + (Z \otimes \Sigma_w^{-1})\tilde{\boldsymbol{\beta}}\}$ and covariance matrix $\Sigma_{\boldsymbol{\alpha}}^* = [S_\alpha^{-1} + (ZZ^{'} \otimes \Sigma_w^{-1})]^{-1}$.

If we do not impose the stability restriction on the VAR(p) process, (19) can be directly used 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 $N(\boldsymbol{\alpha}^*, \Sigma_{\boldsymbol{\alpha}}^*)I(\boldsymbol{\alpha} \in B)$, 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 $N(\boldsymbol{\alpha}^*, \Sigma_{\boldsymbol{\alpha}}^*)$ 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 *et al.* 1996).

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

(a) Define a horizontal slice, $S^h = \{\alpha_i : z < f(\alpha_i)\}$, where z is an auxiliary variable sampled uniformly from $(0, f(\alpha_i^{\text{pre}}))$.

(b) Find an interval, $I = (L, R)$, around α_i^{pre} on S^h such that $f(L) < z$ and $f(R) < z$.

(c) Accept α_i^{new}, sampled uniformly from I, if $f(\alpha_i^{\text{new}}) > z$.

Roberts and Rosenthal (1999) show that the slice sampler is irreducible, 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, since 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}. The 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 the computation of lower and upper bounds of the region B. One need simply check whether or not α_i falls inside the region B, by using (6).[1]

3.2.4 Sampling from $p(\Sigma_w|\beta, b, d, A, \Sigma_b, y)$ and $p(\Sigma_b|\beta, b, d, A, \Sigma_w, y)$

The conditional posterior density for Σ_w is

$$\begin{aligned} p(\Sigma_w|\beta, b, d, A, \Sigma_b, y) &= p(\Sigma_w|\beta, \{\beta\}^0_{m=1-p}, d, A, y) \\ &\propto IW(v^*_w, S^*_w), \end{aligned}$$

an inverted Wishart density with $v^*_w = v_w + T$ and $S^*_w = S_w + \sum_{t=1}^T l_t l_t^{'}$, where $l_t = \beta_t - d - AZ_{t-1}$.

The conditional posterior density for Σ_b is

$$\begin{aligned} p(\Sigma_b|b) &= p(\Sigma_b|y, \beta, b, d, A, \Sigma_w, y) \\ &\propto IW(v^*_b, S^*_b), \end{aligned} \tag{20}$$

an inverted Wishart density with $v^*_b = v_b + H$ and $S^*_b = S_b + \sum_{h \in \mathcal{H}} b_h b_h^{'}$.

3.3 Comparative Model Specifications

The model (2) has parameters $\{\beta_t\}_{t=1}^T$, $\{b_h\}_{h=1}^H$, d, $\{A_n\}_{n=1}^p$, Σ_w and Σ_b. We consider several alternative models that differ by the structural assumptions imposed on d, $\{A_n\}_{n=1}^p$, and Σ_w:

[1]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 $\bar{\boldsymbol{\alpha}} = (\alpha_1^{'}; \alpha_2^{'})^{'}$, with $\alpha_2 = a$, where a is a vector of restricted values. Define a partition matrix P such that $\bar{\boldsymbol{\alpha}} = P\boldsymbol{\alpha}$. Then, the conditional posterior density of α_1 given $\alpha_2 = a$, $N(\alpha_1|\alpha_2 = a)$, can be easily obtained from $N(P\boldsymbol{\alpha}^*, P\Sigma^*_{\boldsymbol{\alpha}}P^{'})$. If the partitioned sub-matrix for α_2 in $P\Sigma^*_{\boldsymbol{\alpha}}P^{'}$ is singular, the Moore-Penrose inverse can be used to derive $N(\alpha_1|\alpha_2 = a)$.

Model	d	$\{A_n\}_{n=1}^p$	Σ_w
No Parameter Dynamics			
M_0 : Static random effects logit model	NR*	0	0
Parameter Dynamics			
M_1 : Dynamic Linear model; random-walk	0	$p = 1; A_1 = I$	NR
M_2 : Random-walk with a drift	NR	$p = 1; A_1 = I$	NR
M_3 : VAR(p)	NR	NR	NR
M_4 : Restricted VAR(p); RVAR(p)	NR	A_n =diagonal	NR

* NR = No Restriction

All models listed in the above table incorporate heterogeneity as a random effects specification; see (7). Model M_0 is the traditional random effects logit, which assumes that there are no parameter dynamics; M_1 to 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 β_t with mean vector β_{t-1} and covariance matrix Σ_w. Model M_2 assumes a random-walk process with a drift term for β_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 $\{A_n\}$ 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 β_t, Σ_w, and $vec(A)$. M_1 can be estimated by skipping the MCMC steps of d and $vec(A)$. Similarly, M_2 can be estimated by skipping the MCMC step for $vec(A)$. 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, differing in the relative complexity of β_t's dynamics. All model parameters were recovered well in each (see Appendix A; 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 training sample to estimate the model and the remaining 6 weeks' data were used for the prediction of the future model parameters. The numbers of observations for the training and future parameter forecasting samples were 6364 and 318, respectively. To ensure identifiability, the time-varying common effect of the fourth option as well as its random effect were fixed to be zero. This requires that the values x_{hjt} for option j be the *differences* of the corresponding predictor variable values for options j and the base option, J. The vector 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, while the last is continuous.

All mean vectors of the normal priors, that is, m_0, m_d, and m_α in (8) to (10), were set to zero. The chosen values for S_0 and S_d were $100I$. For the inverse Wishart priors of Σ_w and Σ_b, the degree of freedom parameters were set to be 2 and the scale parameters were chosen so as to make the expected values $100I$. For the value of S_α in (10) and (13), Litterman (1986) suggested choosing λ, θ, and $\{\sigma_i\}$ by examining the data and by trying several different values. However, this approach entails double-usage of data. We thus set $\lambda = 1.0$, θ to 0.5 and all ratios σ_r/σ_c to 1. For VAR(p) when $p > 1$, S_α are set by using (13) with $\lambda = 1.0$, θ =0.5, and $\frac{\sigma_r}{\sigma_c} = 1$.

4.2 MCMC 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 are smaller, the acceptance rate increases, but we will need a longer chain because the distance between the previous value and a newly accepted value becomes smaller. The chosen tuning constants for β_t and b_h were approximately $\phi_\beta = 0.07$ and $\phi_b = 0.3$, respectively. For all models, the acceptance rates for β_t and b, given these tuning constants, ranged from 53.2% to 55.7% and from 53.8% to 57.5%, respectively.

The number of quantities of interest for the VAR(p) model is quite large. For example, excluding 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 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 β_1 for the first 40,000 iterations, where G represents Geweke's convergence diagnostic with 20,000 burn-in periods. Across all models,

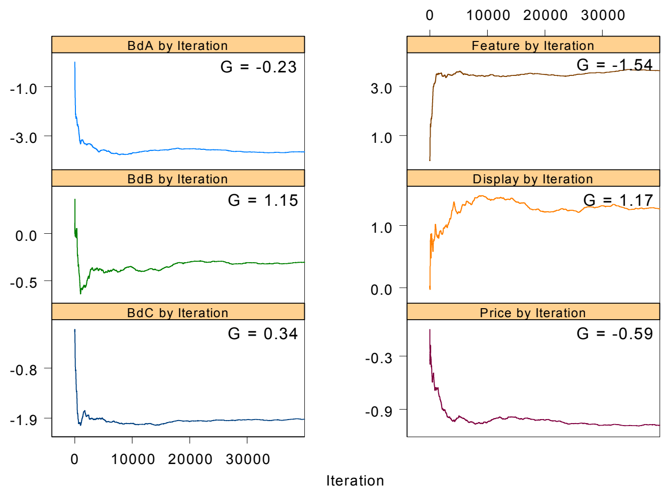


Figure 1: Running mean plot of β_1 of the full VAR(1) model

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 pass. All inferences made here are based on the next 20,000 iterations.

4.3 Tests for Parameter Dynamics

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

$$H_1 \quad : \quad \text{Parameters are static } (M_0)$$

$$H_2 \quad : \quad \text{Parameters display some form of dynamics } (M_1, M_2, M_3, \text{ or } M_4)$$

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, since 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 to M_4, are decisively preferred to the traditional

M_i	Log of Integrated Likelihood	Bayes Factor (BF_{M_0,M_i})
No Parameter Dynamics Case		
M_0 : Static random effects logit model	-3436.33	1.0
Parameter Dynamics Case		
M_1 : Dynamic Linear model	-3068.50	$1.79e-160$
M_2 : Random-walk with a drift	-3072.44	$9.22e-159$
M_3 : VAR(p)		
VAR(1)	-3061.90	$2.44e-163$
VAR(2)	-3075.66	$2.31e-157$
M_4 :RVAR(p)		
RVAR(1)	-3038.89	$2.48e-173$
RVAR(2)	-3063.15	$8.51e-163$

Table 1: Model Comparison for Training Sample

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 parameter dynamics models (M_1 to 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 $= 0.98e + 10$), 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 below, 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, while the "prediction data set" 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 data set, we computed the Bayes factor of M_0 vs. RVAR(1). Results appear in Table 2. Regardless of the value for w, model RVAR(1) is decisively preferred to M_0.

The results for the prediction data set 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 data set (given b_h and β simulated at each MCMC iteration) and taking an average of these choice probabilities for the prediction data set across MCMC iterations. For Model RVAR(1), we did the same given each value of b_h and β_t, $t = w, ..., 96$.

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

	Calibration Sample			Log-Likelihood for Prediction Sample	
Data used	log(Integrated Likelihood) M_0	log(Integrated Likelihood) RVAR(1)	Bayes Factor (H1: M_0)	M_0	RVAR(1)
Weeks 1 − 50	−2048.60	−1782.73	$3.42e-116$	−2256.17	−2236.09
Weeks 1 − 55	−2247.86	−1946.93	$1.33e-123$	−1942.92	−1916.08
Weeks 1 − 60	−2449.33	−2166.87	$2.13e-123$	−1623.90	−1613.33
Weeks 1 − 65	−2571.30	−2315.47	$7.84e-112$	−1396.55	−1388.50
Weeks 1 − 70	−2812.06	−2516.53	$4.50e-129$	−1080.21	−1071.03

Table 2: Cross-validation Comparisons

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 β_{ht} :

$$\begin{aligned} \beta_{ht} &= \beta_t + b_h, \\ \beta_t &= d + A_1\beta_{t-1} + w_t, \\ b_h &\sim N(0, \Sigma_b),\ w_t \sim N(0, \Sigma_w), \end{aligned}$$

where A_1 is a diagonal matrix. The parameters, apart from β_t and b_h, are thus d, diag(A_1), Σ_w, and Σ_b.

An important derived parameter is the long-run mean of β_t, $\mu_\beta = (I - A_1)^{-1}d$. Furthermore, the long-run variance for the i-th element of β_t, is

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

where $\Sigma_{w,i,i}$ and $A_{1,i,i}$ are the i-th diagonal elements of Σ_w and A_1, respectively (see Hamilton 1994 for derivation of the moments of the full VAR(p) process). An indication of the overall variability for element i of β_{ht} is thus given by

$$\mathrm{Var}(\beta_{ht,i}) = \frac{\Sigma_{w,i,i}}{1 - A_{1,i,i}^2} + \Sigma_{b,i,i}\ , \tag{22}$$

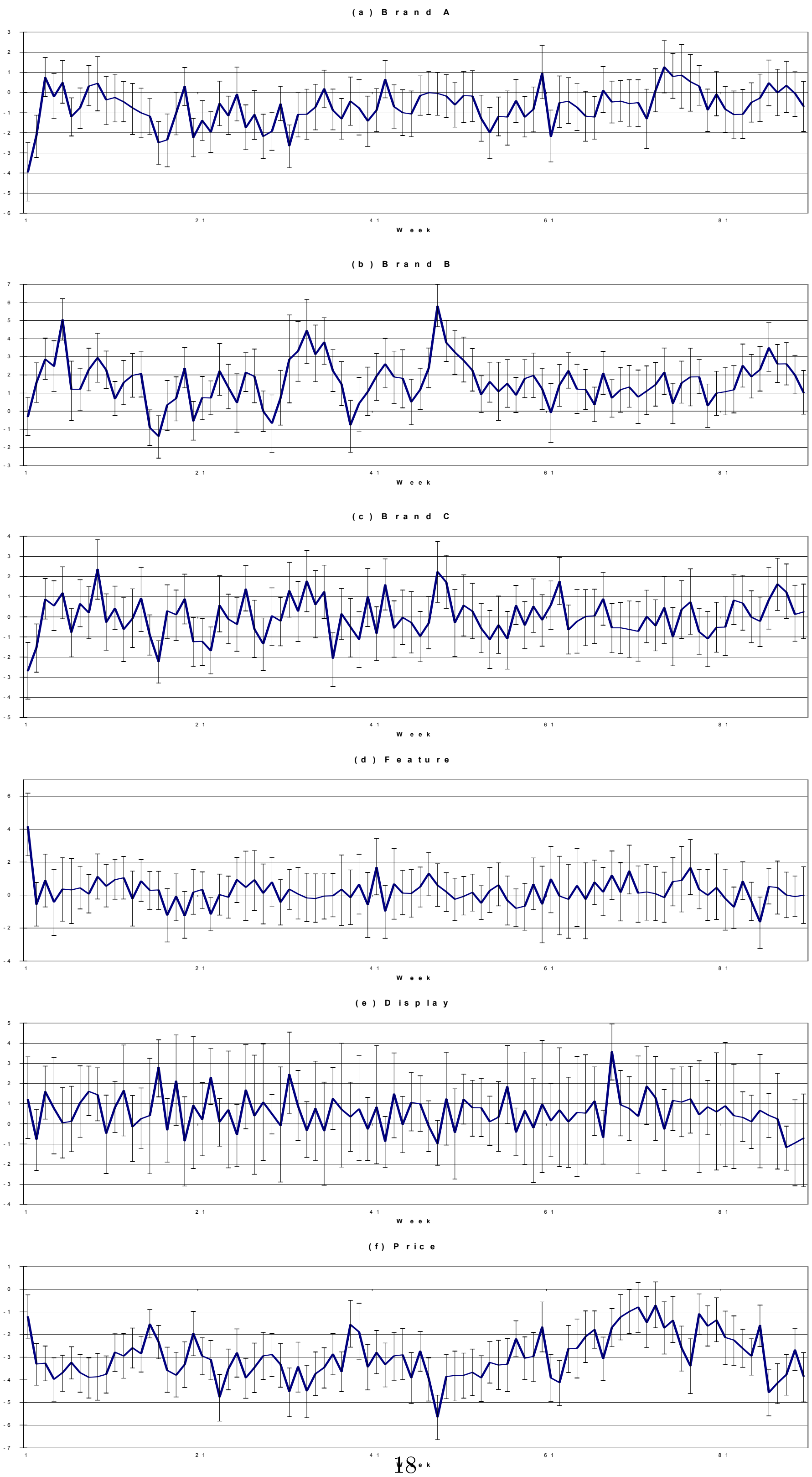
which can be used to give an idea of the relative contributions of parameter dynamics and heterogeneity; it can also be compared with the elements of $E(\beta_{ht}) = \mu_\beta = (I - A_1)^{-1}d$ to get an idea of the relative variability of the elements of β_{ht}.

Estimation of β_t : Figure 2 plots posterior means, fifth and ninety-fifth percentiles for all time-varying parameters β_t. It suggests fairly large temporal fluctuations – all intercepts show strong stochastic patterns, and all variable coefficients display stochastic dynamics. There are several periods which show substantial shifts from β_{t-1} to β_t.

Estimation of d and A_1 : If we define "significant difference" to mean that a (5 percentile, 95 percentile) interval does not contain zero, Table 3 suggests that all elements of d, except the coefficients of the dummy for option C (dum_C) and of Feature, are significantly different from zero. Likewise, the elements of 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 β_t.

		Estimate (std. dev; MC error)	(5 percentile, 95 percentile) interval
d			
	dum_A	-0.5410 (0.2265;0.0094)	(-0.9182,-0.1744)
	dum_B	1.1811 (0.3020;0.0157)	(0.6867, 1.6780)
	dum_C	0.1172 (0.2900;0.0168)	(-0.3634, 0.5982)
	Feature	0.1887 (0.2213;0.0107)	(-0.1722, 0.5536)
	Display	0.6387 (0.3528;0.0182)	(0.1115, 1.2451)
	Price	-2.0741 (0.4470;0.0243)	(-2.7682,-1.3043)
$\mathrm{diag}(A_1)$			
	dum_A	0.2322 (0.1339;0.0051)	(0.0282, 0.4632)
	dum_B	0.2640 (0.1397;0.0058)	(0.0467, 0.4989)
	dum_C	-0.0087 (0.1525;0.0067)	(-0.2328, 0.2655)
	Feature	0.1659 (0.1691;0.0134)	(-0.1411, 0.4213)
	Display	-0.0765 (0.1890;0.0089)	(-0.3999, 0.2241)
	Price	0.2958 (0.1423;0.0068)	(0.0747, 0.5420)

Table 3: Estimates of d and A



Note: 1) Upper and lower bars denote 95th and 5th percentiles, respectively

2) Solid lines denote estimated values

Figure 2: Dynamics of β_t

Estimation of Σ_w Table 4 gives estimates for Σ_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 Σ_w, the ratios of posterior means to posterior standard deviations are between 4.8 and 6.1. Thus, these elements of Σ_w differ from zero, implying that all elements of β_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 data set, β_t is apparently time-varying, with a fairly pronounced degree of white noise.

	dum_A	dum_B	dum_C	Feature	Display	Price
dum_A	2.2014	0.1842	0.1582	-0.0107	-0.0581	0.0797
	(0.3718)					
dum_B	0.4612	2.8483	0.3099	-0.0106	-0.1046	-0.2411
	(0.3033)	(0.5006)				
dum_C	0.3771	0.8401	2.5797	0.0145	-0.0876	-0.2210
	(0.2864)	(0.3567)	(0.4453)			
Feature	-0.0249	-0.0280	0.0364	2.4485	0.0129	0.0094
	(0.2816)	(0.3243)	(0.3059)	(0.4502)		
Display	-0.1580	-0.3238	-0.2581	0.0371	3.3642	0.0715
	(0.3433)	(0.3942)	(0.3784)	(0.3661)	(0.6998)	
Price	0.1756	-0.6038	-0.5267	0.0218	0.2000	2.2031
	(0.2596)	(0.3089)	(0.2931)	(0.2783)	(0.3354)	(0.3630)

Table 4: Estimate of Σ_w

Estimation of Σ_b Table 5 gives estimates for Σ_b. Posterior means are again 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. This table suggests that Σ_b is neither a null matrix nor a diagonal matrix; furthermore, all diagonal elements are significantly different from zero.

Let us briefly examine the effect of parameter dynamics on the heterogeneity distribution by comparing posterior means for the covariance Σ_b for both the RVAR(1) model and model M_0. For model M_0, the posterior mean for the diagonal of Σ_b is

$$\begin{bmatrix} 5.4701 & 8.4474 & 11.1818 & 1.8029 & 1.9395 & 5.3011 \\ (0.6615) & (0.9838) & (1.4515) & (0.2342) & (0.2695) & (0.6255) \end{bmatrix},$$

where the respective posterior standard deviations are given in parentheses. These diagonal elements are 18.2% to 32.0% smaller than the corresponding elements of Σ_b for the RVAR(1) model, which were given in Table 5. Therefore, the traditional random effects logit model 'underestimated' the extent of heterogeneity.

	dum_A	dum_B	dum_C	Feature	Display	Price
dum_A	7.2242	0.1683	0.2161	-0.1588	-0.0699	-0.0151
	(0.8232)					
dum_B	1.5397	11.5826	0.2263	-0.0486	-0.1414	-0.4943
	(0.7671)	(1.3067)				
dum_C	2.3549	3.1222	16.4344	-0.0314	-0.1346	-0.2889
	(0.9208)	(1.1404)	(2.0409)			
Feature	-0.6407	-0.2483	-0.1913	2.2548	-0.0736	0.0311
	(0.4182)	(0.5370)	(0.6885)	(0.3371)		
Display	-0.2898	-0.7423	-0.8419	-0.1706	2.3795	0.1282
	(0.4220)	(0.5843)	(0.7388)	(0.2292)	(0.3906)	
Price	-0.1034	-4.2839	-2.9822	0.1189	0.5036	6.4841
	(0.5319)	(0.7081)	(0.7751)	(0.3935)	(0.4069)	(0.7261)

Table 5: Estimate of Σ_b

Discussion The elements of d and A_1 for the option C dummy are essentially 0, but the corresponding variance in Σ_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. Since 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-zero mean, while the Price coefficient appears to follow a AR(1) process, over the observation period.

Table 6 gives the posterior means for the following quantities:

- μ_β, the long-run mean of β_{ht} (and the posterior standard deviation of μ_β),
- $\sqrt{\text{Var}(\beta_{ht,i})} = \sqrt{\frac{\Sigma_{w,i,i}}{1-A_{1,i,i}^2} + \Sigma_{b,i,i}}$, an overall standard deviation for β_{ht},
- $\sqrt{\Sigma_{b,i,i}}$, the standard deviation of the heterogeneity component, b_h, of β_{ht},
- $\sigma_{\beta,i} = \sqrt{\frac{\Sigma_{w,i,i}}{1-A_{1,i,i}^2}}$, the long-run standard deviation of the dynamic component, β_t, of β_{ht}, and
- $\sqrt{\Sigma_{w,i,i}}$, the standard deviation of the "white noise" component of β_t.

The posterior mean of μ_β 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 μ_β. A comparison of the results for μ_β with those for d in Table 3 shows a moderate difference for Price.

The overall variability in β_{ht}, as measured by the posterior mean for the standard deviation $\sqrt{\text{Var}(\beta_{ht,i})}$, is quite large. In fact, all these standard deviations are *larger* than the corresponding elements of μ_β, so the corresponding regression coefficients are negative for some households and time periods, and positive for others. Thus, though the posterior means for the 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 Σ_b, $\sqrt{\Sigma_{b,i,i}}$, and parameter dynamics by $\sigma_{\beta,i} = \sqrt{\Sigma_{w,i,i}/(1 - A_{1,i,i}^2)}$; see (21). Except for Feature and Display, Table 6 suggests that 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, $\beta_t = d + A_1\beta_{t-1} + w$, with $\sqrt{\Sigma_{w,i,i}}$, the standard deviation of the "white noise" component, w_t. The value of the posterior mean for $\sqrt{\Sigma_{w,i,i}}$ 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 β_t.

i	μ_β	$\sqrt{\text{Var}(\beta_{ht,i})}$	$\sqrt{\Sigma_{b,i,i}}$	$\sigma_{\beta,i}$	$\sqrt{\Sigma_{w,i,i}}$
dum_A	-0.7037	3.0965	2.6834	1.5385	1.4786
	(0.2739)	(0.1523)	(0.1528)	(0.1433)	(0.1235)
dum_B	1.6095	3.8334	3.3979	1.7667	1.6813
	(0.3129)	(0.1910)	(0.1920)	(0.1664)	(0.1462)
dum_C	0.1147	4.3613	4.0462	1.6201	1.6003
	(0.2882)	(0.2407)	(0.2499)	(0.1415)	(0.1365)
Feature	0.2270	2.2001	1.4975	1.6067	1.5583
	(0.2710)	(0.1423)	(0.1115)	(0.1551)	(0.1416)
Display	0.5849	2.4236	1.5375	1.8671	1.8246
	(0.2909)	(0.1809)	(0.1247)	(0.2023)	(0.1873)
Price	-2.9491	2.9930	2.5424	1.5728	1.4794
	(0.2833)	(0.1504)	(0.1423)	(0.1518)	(0.1206)

Table 6: Elements of variation in β_{ht}

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 data into nine data, sets such that $\bar{y}_z = \{y_t\}_{t=10(z-1)+1}^{10z}$, where $z = 1, \ldots, 9$, we estimate all nine regression coefficients $\{\bar{\beta}_z\}_{z=1}^{9}$ for these nine data sets 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:

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

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

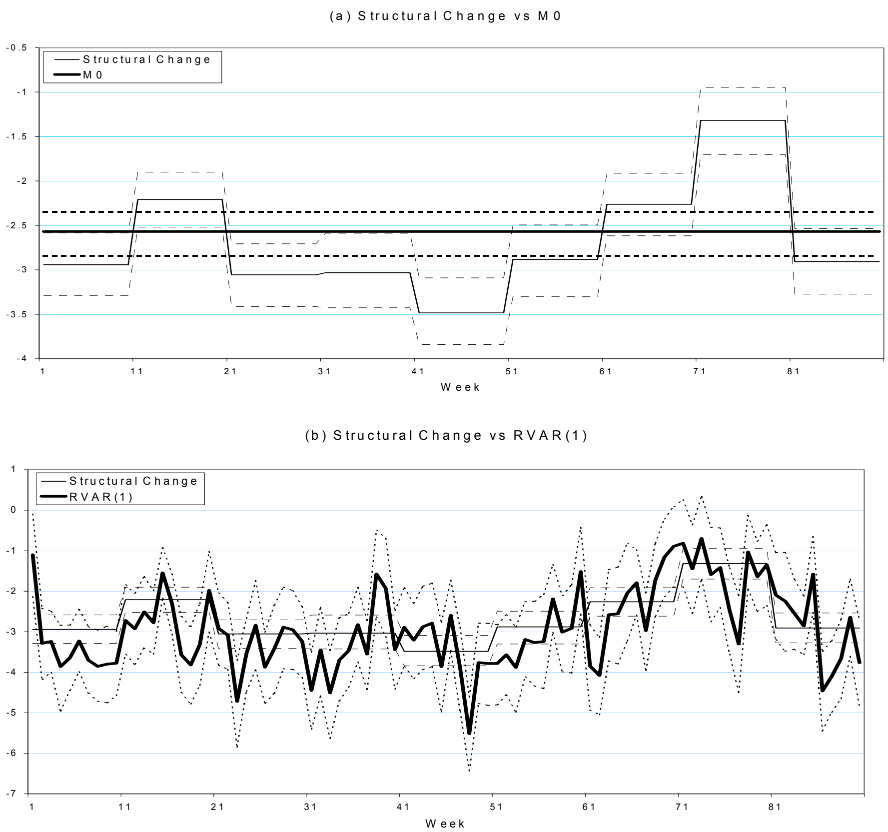
$$\begin{aligned} \text{H}_1 \quad &: \quad \bar{\beta}_1 = ... = \bar{\beta}_9 \ (M_0) \\ \text{H}_2 \quad &: \quad \bar{\beta}_1 \neq ... \neq \bar{\beta}_9 \ \text{(Structural change model)} \end{aligned}$$

The log of the integrated likelihood of the structural change model is -3351.46. Clearly, the null hypothesis H_1 is rejected (Bayes Factor favoring H_1 over H_2 =$1.38e-37$). This in turn further verifies that there do exist parameter dynamics for these data. Note that the RVAR(1) model is still decisively preferred to the structural change model (Bayes Factor favoring RVAR(1) over the structural change model = $5.59e+135$), which implies 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, 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 both those of 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.



Note: The lower and upper dotted lines enveloping each solid line of estimates denote the corresponding 5th and 95th percentiles, respectively.

Figure 3: Comparison of price coefficient estimates for the structural change and RVAR(1) models

4.8 Effects of Parameter Dynamics on Choice Behavior

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

From (2), define the following derivatives:

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

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 above 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 k-th 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 shows the MCMC estimates of these household-averaged derivatives further averaged over the 90 week observation period, e.g., $\bar{\rho}_{x_{t,k}}(j,i) = \frac{1}{90}\sum_{t=1}^{90}\rho_{x_{t,k}}(j,i)$. The pattern of results indicates that M_0 tends to overestimate all quantities of interest..

	M_0	RVAR(1)
$\bar{\rho}_{x_{t,k}}$(A,A)	-0.1515 (0.0047)	-0.1408 (0.0052)
$\bar{\rho}_{x_{t,k}}$(A,B)	0.0616 (0.0028)	0.0556 (0.0030)
$\bar{\rho}_{x_{t,k}}$(A,C)	0.0333 (0.0021)	0.0324 (0.0023)
$\bar{\rho}_{x_{t,k}}$(A,D)	0.0565 (0.0027)	0.0528 (0.0028)
$\bar{\rho}_{\beta_{t,k}}$(A)	-0.0461 (0.0015)	-0.0383 (0.0013)
$\bar{\rho}_{\beta_{t,k}}$(B)	0.0404 (0.0015)	0.0327 (0.0013)
$\bar{\rho}_{\beta_{t,k}}$(C)	0.0227 (0.0012)	0.0191 (0.0011)
$\bar{\rho}_{\beta_{t,k}}$(D)	-0.0170 (0.0013)	-0.0135 (0.0011)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}$(A,A)	0.0559 (0.0023)	0.0447 (0.0022)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}$(A,B)	-0.0151 (0.0014)	-0.0119 (0.0013)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}$(A,C)	-0.0110 (0.0012)	-0.0087 (0.0011)
$\bar{\rho}_{x_{t,k}\beta_{t,k}}$(A,D)	-0.0298 (0.0011)	-0.0241 (0.0011)

Note: Standard deviations appear in parentheses

Table 7: The averaged effects of covariate and parameters on choice behavior

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 β_{T+m}.

We conducted 6-step ahead forecasting, obtaining posterior distributions for β_{T+m}, where $T = 90$ and $m = 1, \ldots, 6$. At each MCMC iteration, it is straightforward to simulate these future parameters, β_{T+m}, given β, d, A_1, and Σ_w by using (5). The posterior means and standard deviations of β_{T+m} are, therefore, readily available from MCMC runs of β_{T+m}. Their means and standard deviations for weeks 91 to 96 are given in Table 8, suggesting considerable forecasting uncertainty.

	dum_A	dum_B	dum_C	Feature	Display	Price
Week 91	-0.6915	1.4296	0.1153	0.1756	0.6704	-3.1938
	(0.2841)	(0.3402)	(0.3081)	(0.3335)	(0.4910)	(0.3367)
Week 92	-0.7000	1.5482	0.1206	0.2119	0.5336	-3.0395
	(0.2620)	(0.3015)	(0.2883)	(0.2686)	(0.3057)	(0.2871)
Week 93	-0.7024	1.5858	0.1151	0.2223	0.5946	-2.9868
	(0.2671)	(0.3037)	(0.2880)	(0.2672)	(0.3020)	(0.2797)
Week 94	-0.7032	1.5993	0.1151	0.2252	0.5790	-2.9664
	(0.2705)	(0.3074)	(0.2881)	(0.2686)	(0.2895)	(0.2798)
Week 95	-0.7035	1.6048	0.1147	0.2263	0.5868	-2.9576
	(0.2721)	(0.3097)	(0.2881)	(0.2698)	(0.2930)	(0.2808)
Week 96	-0.7036	1.6072	0.1147	0.2267	0.5837	-2.9535
	(0.2729)	(0.3110)	(0.2881)	(0.2703)	(0.2901)	(0.2817)
Long-term mean	-0.7037	1.6095	0.1147	0.2270	0.5849	-2.9491

Note: Standard deviations appear in parentheses

Table 8: 6-step ahead forecasting of β_t

We compared the performance of the six forecasted β_{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 data set. These likelihood values were obtained by first computing predicted choice probabilities, (2), for the forecasting data set given all parameters simulated at each MCMC iteration and 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 and -208.89 for RVAR(1) and M_0, respectively. For the future parameter forecasting sample, RVAR(1) shows 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 vector autoregressive framework to account for the phenomenon, one which can be grafted onto any specifications for utility or error structure. In this framework, one 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, though not all, of the parameters showed strong evidence of temporal variation. This was clear even under the parsimonious specification which 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 it merits study in and of itself, given the popularity of the random effects logit modeling framework. Finally, our data suggested that choice dynamics may be mis-attributed to exogenous covariates when parameters are presumed not to have dynamics of their own. For example, the random effects model appears to under-adjust 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 showed that, if all or some of the population update their parameters over time, 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) note that scanner panel data do not accommodate the proper unit of analysis in modeling preference changes: a person-activity occasion. They discuss how, for many activities (e.g., snacking, serving wine), the consumer environment isn't constant from one usage occasion to another, so that preferences are rightly situationally or motivationally dependent. Although they explicitly point out that occasions for use of laundry detergent, the product class used in our study, are less likely to be subject to this sort of temporal preference variation, we believe 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 approaches to this practical problem are presented at length in their paper.

In a similar vein, Wakefield and Inman (2003) also note that little research has focused on effects of consumption occasion or context on consumer price sensitivity. They found price sensitivity to be attenuated by hedonic and social consumption situations: since intended consumption occasion varies across consumers and time, that this variation is unobserved 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 a purely seasonal explanation could support. We believe that such issues could be directly addressed by 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, both in terms of data requirements and estimation technology, though we suspect each of these impediments will wane with time.

The model is not without its limitations. One such limitation is the requirement for data over a relatively long period. In many applications, particularly so 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 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 applied widely in choice research, due both to its generality and its silence on utility and error structure. We believe it can be extended readily 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.

References

[1] Allenby, G. M. and P. J. Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of American Statistical Association*, 89, 428, 1218-1231.

[2] ______ and ______ (1995), "Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice," *Journal of Business & Economic Statistics*, 13, 3, 281-289.

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

[4] Bernardo, J. M. and A. F. M. Smith (1994), *Bayesian Theory*, New York: Jon Wiley and Sons Ltd.

[5] Cargnoni, C., P. Müller, and M. West (1997), "Bayesian Forecasting of Multinomial Time series Through Conditionally Gaussian Dynamic Models," *Journal of the American Statistical Association*, 92, 438, 640-647.

[6] Carlin, B. P., N. G. Polson, and D. S. Stoffer (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87 (418), 493-500.

[7] Chib, S. and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithms," *The American Statistician*, 49, 327-335.

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

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

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

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

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

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

[14] Kahneman, D. and A. Tversky, A. (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 2, 263-291.

[15] Kass, R. E. and A. E. Raftery (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 430, 773-795.

[16] Li, H. and R. S. Tsay (1998), "A Unified Approach to Identifying Multivariate Time Series Models," *Journal of the American Statistical Association*, 93, 442, 770-782.

[17] Litterman, R. B. (1986), "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience," *Journal of Business & Economic Statistics*," 4, 25-38.

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

[19] Manski, C. F. (1977), "The structure of Random Utility Models," *Theory and Decision*, 8, 229-254.

[20] McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka ed. *Frontiers in Econometrics*, New York: Academic.

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

[22] Neal, R. M. (1997), "Markov Chain Monte Carlo Methods Based on 'Slicing' the Density Function," *Technical Report*, No. 9722, Department of Statistics, University of Toronto.

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

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

[25] Roberts, G. O. and J. S. Rosenthal (1999), "Convergence of Slice Sampler Markov Chains," *Journal of the Royal Statistical Society: Series B*, 61, 643-660.

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

[27] Thaler, R. (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4, 3, 199-214.

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

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

[30] Wakefield, Kirk L. and J. Jeffrey Inman (2003), "Situational Price Sensitivity: The Role of Consumption Occasion, Social Context, and Income," *Journal of Retailing*, 79(4), 199-212.

[31] West, M. and P. Harrison (1997), Bayesian Forecasting and Dynamic Models, New York: Springer, 2nd ed.

[32] Yang, Sha, Greg M. Allenby and Geraldine Fennell (2002) "Modeling Variation in Brand Preference: The Roles of Objective Environment and Motivating Conditions," *Marketing Science*, 21, 1, 14-31.

A Appendix: Estimation Based on Synthetic Data

The model outlined in the previous sections contains many parameters and the estimation procedure is complex. This appendix provides some results about the ability of our model to recover its parameters. To this end, we conduct a simulation study using synthetically generated data based on known parameters.

A.1 Simulation

We generated a synthetic data set with relatively complex dynamics of β_t. We take $J = 3$, $T = 50$, and $k = 5$. The first two elements in the predictor variables were dummy variables for option A and B, $x_{hjt,1}$ and $x_{hjt,2}$ The other three predictor variables, $x_{hjt,3}$, $x_{hjt,4}$, and $x_{hjt,5}$ were randomly generated from Ber(0.6), Ber(0.4), and $n(x_{hjt,5}|2{,}3)I_{1<x_{hjt,5}<3}$, respectively.

We used a VAR(1) process for generating the data on β_t, where the known parameter values were

$$d = \begin{pmatrix} 0.5 \\ -0.5 \\ 1.0 \\ 1.5 \\ -3.0 \end{pmatrix}, A_1 = \begin{pmatrix} .5 & 0 & .3 & 0 & 0 \\ 0 & .5 & 0 & 0 & .3 \\ 0 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 \end{pmatrix}, \text{ and } \beta_0 = \begin{pmatrix} 2.1 \\ -3.6 \\ 2.7 \\ 1.8 \\ -4.9 \end{pmatrix}.$$

In addition, we chose Σ_w and Σ_b such that their diagonal elements were 2 and 3, respectively; the off-diagonal elements of both matrices were set to be 0.1. The long-term mean of β_t was thus

$$\mu_\beta = (I - A_1)^{-1} d = \begin{pmatrix} 2.2 \\ -4.6 \\ 2.0 \\ 1.5 \\ -6.0 \end{pmatrix}$$

To obtain the synthetic data set, as before we first generated data for 200 individuals. Each week, 100 of these individuals were randomly chosen to make choices. We removed individuals for which either $x_{hjt,3}$ or $x_{hjt,4}$ has the same value for all chosen options across their choice occasions. Our final artificial data set contained $H = 165$ individuals. The total number of choice observations for the simulated data was 4115 and the weekly sample size ranged from 78 to 87. The total number of choice occasions across individuals ranged from 19 to 31.

The values of the prior distribution for the parameters, β_0, d, A, Σ_w, and Σ_b, given in (8) to (12), were $m_0 = m_d = m_\alpha = 0$, $v_w = v_b = 2$, $S_0 = S_d = S_w = S_b = 30I$. In addition, for S_α in (10) and (13), we set $\lambda = 1.0$ to ensure unity variances for all diagonal elements of A_1 because all diagonal elements of A_1 under the stability condition must have values between -1 and 1. The chosen value of θ is 0.5 and we set all ratios σ_r/σ_c to be 1.

First, we fitted M_0 and VAR(p) models to the synthetic data set and computed Bayes factors, (15). To ensure identifiability, the time-varying common effect of the fourth option as well as its random effect were fixed to be zero. The MCMC was run for 40,000 iterations, and it appears to have converged after 20,000 iterations for all models. Note that the assumed true parameter dynamics is the VAR(1) process. The computed Bayes factors are given in Table 9. As shown in Table 9, VAR(1) was preferred to M_0 and other VAR(p) models. This result confirms that the proposed model can recover the underlying true parameter dynamics pretty well.

	$\log(\text{BF}_{M_0,i})$	$\log(\text{BF}_{VAR(1),i})$
M_0	0	1752.44
VAR(1)	−1752.44	0
VAR(2)	−1737.73	14.71
Restricted VAR(1)	−944.16	808.28
Restricted VAR(2)	−924.45	827.97

Table 9: Model comparison on synthetic data set II

Since the most preferred model is VAR(1), we present estimation results for that model. We kept the 20,000 last iterations to estimate the parameters. The MCMC sampler recovers the parameters for the VAR(1) model fairly well (See Tables 11, 10 and Figure 4). Since all off-diagonal elements of estimated Σ_w and Σ_b had 0 in a 95% posterior interval, we do not report estimates of the off-diagonal elements of Σ_w and Σ_b. Two off-diagonal non-zero elements in A_1, $A_{1,13}$ and $A_{1,25}$, were recovered well as well. The estimates of these two quantities were 0.0619 (0.1439) and 0.2516 (0.1191) for $A_{1,13}$ and $A_{1,25}$, respectively, where the posterior standard deviations are given in parenthesis. All other off-diagonal elements in A_1 have 0 within the 95% posterior interval. Last, the long-term mean of β_t was also well recovered. The posterior mean of the long-term mean of β_t was [1.6906 -4.5278 1.4430 1.7430 -5.8324] with posterior standard deviation [9.3770 2.6952 11.0165 4.2708 11.7869].

k	1	2	3	4	5
β_0	2.2387	-5.2161	1.8536	4.0060	-2.0104
	(3.5475)*[2.1]**	(4.2338)[-3.6]	(3.0860)[2.7]	(3.6684)[1.8]	(2.8223)[-4.9]
d	1.2386	-2.6185	0.7199	1.1014	-3.3214
	(1.1607)[0.5]	(1.1893)[-0.5]	(1.1404)[1.0]	(1.1616)[1.5]	(1.3533)[-3.0]
$diag(A_1)$	0.5672	0.1890	0.5226	-0.1172	0.5312
	(0.1568)[0.5]	(0.1745)[0.5]	(0.1499)[0.5]	(0.1735)[0.0]	(0.1550)[0.5]
$diag(\Sigma_w)$	2.6411	2.5189	2.7204	2.9867	3.1370
	(0.6808)[2]	(0.6855)[2]	(0.7229)[2]	(0.8050)[2]	(0.8996)[2]
$diag(\Sigma_b)$	3.0204	2.5276	3.7144	3.2268	4.3155
	(0.6305)[3]	(0.5979)[3]	(0.8577)[3]	(0.7021)[3]	(1.0744)[3]

Note: *: posterior standard deviations; **: true values

Table 10: Estimates for the synthetic data set II

$$\begin{pmatrix} 0.57(0.16) & -0.05(0.17) & 0.06(0.14) & -0.17(0.16) & 0.08(0.12) \\ 0.02(0.14) & 0.11(0.17) & -0.05(0.14) & 0.06(0.15) & 0.25(0.12) \\ 0.08(0.15) & -0.08(0.16) & 0.52(0.15) & -0.01(0.17) & 0.08(0.12) \\ 0.18(0.15) & -0.05(0.16) & -0.04(0.15) & -0.12(0.17) & -0.04(0.12) \\ 0.18(0.17) & -0.10(0.19) & -0.04(0.17) & -0.04(0.19) & 0.53(0.16) \end{pmatrix}$$

Table 11: Estimated A_1 for synthetic data set II

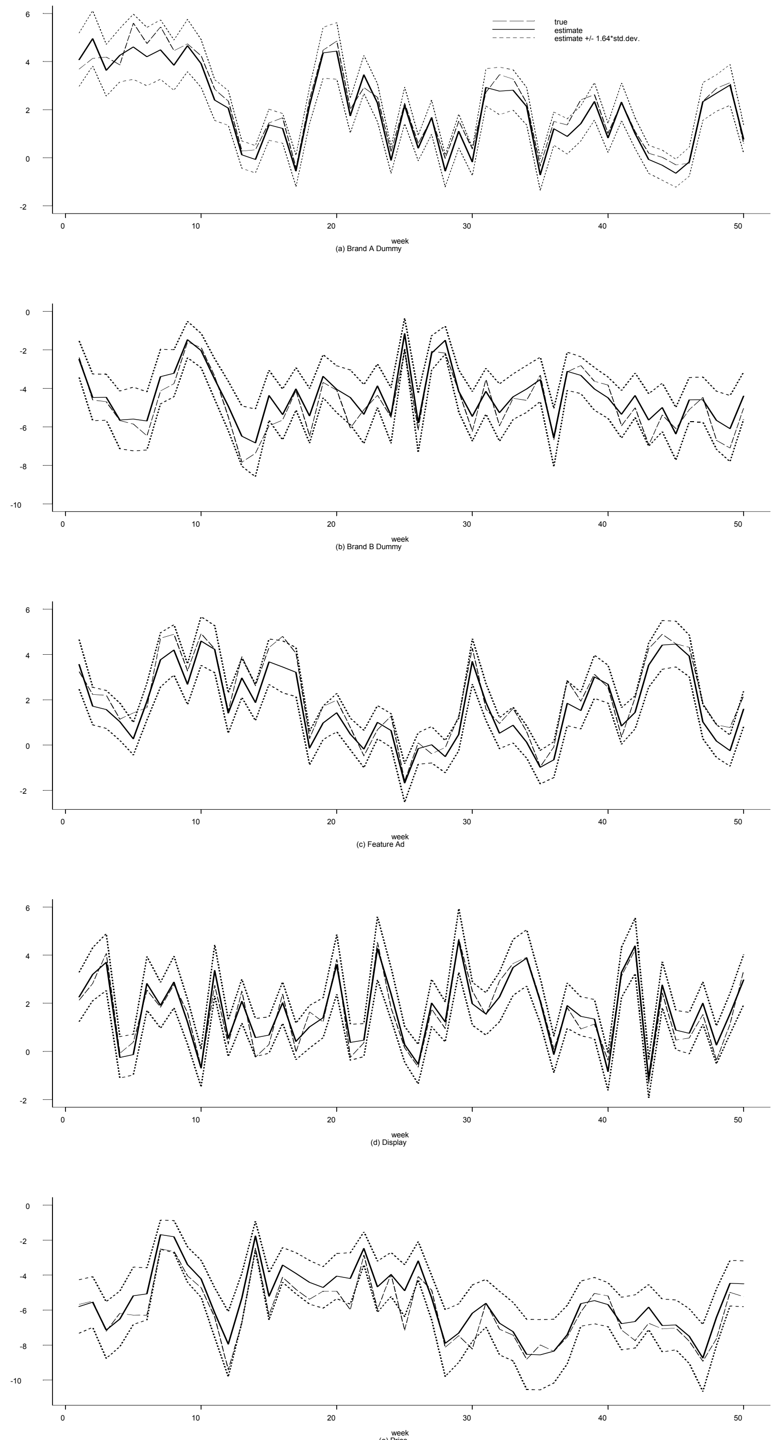


Figure 4: Estimated β_t in the VAR(1) synthetic data set II