Random Coefficient Logit Model for Large Datasets

Carlos Hernández Mireles and Dennis Fok

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2010-021-MKT
Publication	May 2010
Number of pages	46
Persistent paper URL	http://hdl.handle.net/1765/19669
Email address corresponding author	mireles@ese.eur.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES

RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	We present an approach for analyzing market shares and products price elasticities based on large datasets containing aggregate sales data for many products, several markets and for relatively long time periods. We consider the recently proposed Bayesian approach of Jiang et al [Jiang, Renna, Machanda, Puneet and Peter Rossi, 2009. Journal of Econometrics 149 (2) 136-148] and we extend their method in four directions. First, we reduce the dimensionality of the covariance matrix of the random effects by using a factor structure. The dimension reduction can be substantial depending on the number of common factors and the number of products. Second, we parametrize the covariance matrix in terms of correlations and standard deviations, like Barnard et al. [Barnard, John, McCulloch, Robert and Xiao-Li Meng, 2000. Statistica Sinica 10 1281-1311] and we present a Metropolis sampling scheme based on this specification. Third, we allow for long term trends in preferences using time-varying common factors. Inference on these factors is obtained using a simulation smoother for state space time series. Finally, we consider an attractive combination of priors applied to each market and globally to all markets to speed up computation time. The main advantage of this prior specification is that it let us estimate the random coefficients based on all data available. We study both simulated data and a real dataset containing several markets each consisting of 30 to 60 products and our method proves to be promising with immediate practical applicability.
Free Keywords	random coefficient logit, aggregate share models, bayesian analysis
Availability	
Classifications	

Random Coefficient Logit Models for Large Datasets

Carlos Hernández Mireles and Dennis Fok
Erasmus Research Institute of Management
Erasmus University Rotterdam

May 3, 2010

Abstract

We present an approach for analyzing market shares and products price elasticities based on large datasets containing aggregate sales data for many products, several markets and for relatively long time periods. We consider the recently proposed Bayesian approach of Jiang et al [Jiang, Renna, Machanda, Puneet and Peter Rossi, 2009. Journal of Econometrics 149 (2) 136-148] and we extend their method in four directions. First, we reduce the dimensionality of the covariance matrix of the random effects by using a factor structure. The dimension reduction can be substantial depending on the number of common factors and the number of products. Second, we parametrize the covariance matrix in terms of correlations and standard deviations, like Barnard et al. [Barnard, John, McCulloch, Robert and Xiao-Li Meng, 2000. Statistica Sinica 10 1281-1311] and we present a Metropolis sampling scheme based on this specification. Third, we allow for long term trends in preferences using time-varying common factors. Inference on these factors is obtained using a simulation smoother for state space time series. Finally, we consider an attractive combination of priors applied to each market and globally to all markets to speed up computation time. The main advantage of this prior specification is that it let us estimate the random coefficients based on all data available. We study both simulated data and a real dataset containing several markets each consisting of 30 to 60 products and our method proves to be promising with immediate practical applicability.

KEYWORDS: Random Coefficient Logit, Aggregate share models, Bayesian analysis

1 Introduction

A growing number of scholars is developing estimation methods for random coefficient logit models based on aggregate sales data. Currently, the estimation methods are based on the generalized method of moments [GMM], as in Nevo (2001) and Berry et al. (1995) (hereafter BLP), or based on likelihood or Bayesian approaches, as in Jiang et al. (2009) (hereafter Bayesian BLP or BBLP), Yang et al. (2003), and Park and Gupta (2009). The choice of the estimation method depends on the modeling assumptions regarding aggregate demand shocks, consumer heterogeneity, stability of preferences, price endogeneity and on the size and type of data available.

Recent Bayesian and maximum likelihood-based approaches have been successfully applied to data containing relatively long time series of weekly data (ranging from one to six years) concerning a small number of products (usually less than 6 products) sold in a single market (Jiang et al., 2009; Musalem et al., 2006; Yang et al., 2003). The GMM approach has been applied to similar sized data, like in Goeree (2008). Recently, Berry and Pakes (2007) use GMM and apply an extension of the BLP model to data consisting of both small (between 2 and 10) and large (100) number of products. The extension of Berry and Pakes (2007) is mainly focused on relaxing the assumption of non-zero demand shocks specifically when the market is saturated with many products. Their specification of null demand shocks may decrease the precision of the BLP contraction mapping and they present new complementary routines that overcome this issue.

One of the most challenging aspects for all methods is the estimation of the underlying distribution of the random effects that describe individual level consumer heterogeneity. As only aggregate data is available, the heterogeneity needs to be identified based on switching patterns. The simulation results of Jiang et al. (2009) suggest that their Bayesian method performs well and makes a more efficient use of the data relative to a GMM estimator. Nonetheless, today still little is known about the scalability (that is the performance and adaptability) of current methodologies

to settings with many products and markets.

In this paper we investigate the scalability of the Bayesian approach proposed by Jiang et al. (2009) and we extend their method in four directions. First, we propose a factor structure for the covariance matrix of the random effects. We will assume that the covariance matrix between J products can be modeled by a group of K factors, where the factor loadings are based on observable characteristics. Such a structure helps to keep the dimension of the heterogeneity structure under control. That is, we make the same distributional assumptions as in Jiang et al. (2009) regarding the heterogeneity and aggregate demand shocks but we strongly reduce the dimension of the covariance matrix. This reduction will be especially important in applications with a large number of products.

Second, we specify the covariance matrix following Barnard et al. (2000) as a function of correlations and standard deviations and we propose a Metropolis sampling scheme based on this parametrization. This parametrization has two main advantages. A technical advantage is that splitting the covariance in variances and correlations allows for a more efficient sampling scheme. A practical advantage is that the correlation structure of the heterogeneity itself may be most informative for managers.

The third extension in our approach is that we allow for time variation in preferences. Preference fluctuations are likely to occur over long periods of time and over seasons. In the currently considered Random Coefficient Logit Models such developments are often ignored. One exception we are aware of is Chintagunta et al. (2005), who show that allowing for time variation in preferences is beneficial to reduce both the uncertainty regarding brand preferences and the uncertainty regarding the sensitivity of products' shares to marketing efforts.

Finally, we consider an attractive combination of priors applied to each market and globally to all markets. This prior specification let us analyze all data simultaneously and it facilitates the estimation of the underlying distribution of the random coefficients based in all the data.

The Bayesian approach we use in this paper allows for an efficient implementa-

tion of the four extensions mentioned above. One main advantage of the Bayesian approach over simulated maximum likelihood and GMM is that inference over any function of the parameters is straightforward because we obtain the posterior distribution of all parameters as the MCMC output. This for example allows for a straightforward assessment of the uncertainty in (cross) price elasticities. A second main advantage of the Bayesian approach is that we can incorporate efficient sampling of time-varying parameters alongside the other model parameters. Chintagunta et al. (2005) use MLE and specify brand-specific time-fixed effects to account for time variation in preferences. Their specification of brand and time-specific fixed effects is an attractive formulation but the number of fixed effects to estimate may increase rapidly as the number of brands and time periods increases. As Chintagunta et al. (2005), we allow for time-variation in preferences but we use the simple and efficient simulation smoother of Durbin and Koopman (2002) to sample the time-varying parameters. The smoother is flexible because it let us reduce the model to the setting where brand preferences are fixed in time and this reduction may depend on the model's parameter estimates or it can easily be specified a priori.

We illustrate our approach using both simulated data and a real dataset that contains sales data for more than 20 markets each with a different, large, number of products and brands. The remainder of the paper is structured as follows. In the next section we discuss the model. Next we present the Bayesian inference (some technical details are discussed in the appendix). Section 4 shows the results of a simulation experiment. In Section 5 we show detailed results of the application of the model to actual data. We conclude the paper with a discussion.

2 Augmented Bayesian BLP Model

In this section we present our approach and we discuss how we augment the BBLP model in the directions discussed earlier. First in subsection 2.1 we present the model specification. Next in subsection 2.2 we discuss the share inversion method

and the integration of the share function.

2.1 Model Specification

Consider consumers who make purchases from a set of J products during T time periods in M different markets. In general not all products will be available in all markets. We will use the letter J to refer to the total number of unique products available across all markets. $\mathcal{J}^m$ denotes the set of products that are available in market m. The size of this set, that is, the number of products available in market m is denoted by J^m. In each period a consumer in market m can either choose to purchase one of the products in $\mathcal{J}^m$ or choose an outside good, that is, he buys a product outside the set $\mathcal{J}^m$.

The purchase behavior of individual i in market m is based on utility maximization. We assume that the (latent) utility for consumer i for product j at time t in market m (denoted by u_{ijt}^m) contains three parts, (i) an "explained" part (w_{ijt}^m), (ii) a market level aggregate demand shock (η_{jt}^m), and (iii) an individual level random effect (ϵ_{ijt}^m), that is, we specify

$$u_{ijt}^m = w_{ijt}^m + \eta_{jt}^m + \epsilon_{ijt}^m, \qquad j \in \mathcal{J}^m, t = 1, \ldots, T. \tag{1}$$

We make the standard assumption of a type-I extreme value distribution for ϵ_{ijt}^m and we assume $\eta_{jt}^m \sim N(0, \tau_m^2)$. We use a factor structure to further model w_{ijt}^m, that is we use

$$w_{ijt}^m = f_{it}^{m\prime} \lambda_{jt}^m, \qquad j \in \mathcal{J}^m, t = 1, \ldots, T, \tag{2}$$

where f_{it}^m denotes an individual-specific K^m dimensional dynamic factor, and λ_{jt}^m is a $(K^m \times 1)$ vector containing the factor loadings for product j in market m. The factor loadings are based on observable product characteristics, such as, packaging and brand name, but also (log) price and promotional indicators may be part of the factor loading vector. In general λ_{jt}^m will contain constant as well as time-varying elements. In principle the same factors will be used in all markets, however, in

some cases some factors may not be present in a market. For example, a particular package may not yet be available in a market. Therefore we need to specify the number factors to be dependent on the market.

The factor f_{it}^m gives the importance of a particular product characteristic for individual i in market m at time t. We split this factor into a time-varying part, which is the same across the population, and a heterogeneous part, which is constant over time, that is,

$$f_{it}^m = \bar{f}_t^m + \upsilon_i^m, \text{ where } \upsilon_i^m \sim \phi(0, A^m \Psi A^{m\prime}), \tag{3}$$

where Ψ denotes the variance matrix of all individual level random effects and A^m denotes a selection matrix. This matrix selects the rows and columns of the variance matrix that correspond to factors that are relevant for market m. The matrix A^m can be obtained by deleting all rows from the K dimensional identity matrix that correspond to irrelevant factors. Note that the variance of the random effects is in principle common across markets. Together with the factor loadings in λ_{jt}^m the covariance matrix $A^m \Psi A^{m\prime}$ gives a flexible but parsimonious specification for the variance structure of the preference heterogeneity.

Note that we can write the covariance matrix of the utilities for all products in market m, call this matrix Σ^m, as a function of Ψ, the selection matrices A^m and the factor loadings Λ_t^m where $\Lambda_t^m = \{\lambda_{jt}^m\}_{j \in \mathcal{J}^m}$. That is,

$$\Sigma^m = \Lambda_t^{m\prime} A^{m\prime} \Psi A^m \Lambda_t^m. \tag{4}$$

Next we assume a particular law of motion for $\bar{f}_t^m$, the common dynamic component of the factor. We use the state space specification

$$\bar{f}_{t+1}^m = \Gamma_t^m \bar{f}_t^m + \Pi_t^m \omega_t^m, \tag{5}$$

where $\omega_t^m \sim N(0, \Omega^m)$ and Γ_t^m is a known matrix. In the state space literature, Ω^m and Γ_t^m are usually set to be diagonal. Furthermore, if we additionally restrict the

k-th diagonal element of Γ_t^m to be 1, we obtain a random walk for the k-th factor, that is, $\bar{f}_{kt+1}^m = \bar{f}_{kt}^m + \omega_{kt}^m$. If the variance of ω_{kt}^m is set to zero (or the corresponding element of Π_t^m), we obtain a constant specification for the factor, $\bar{f}_{kt}^m = \bar{f}_{k1}^m$. If we instead set the diagonal element of Γ_t^m to zero and the corresponding variance to a non-zero value, we obtain independent random effects over time, $\bar{f}_{k,t+1}^m = \omega_{kt}^m$.

We complete the model by normalizing the utility of the outside good to be 0. Based on the complete utility specification we can derive the purchase probabilities, or consumption share for individual i, s_{ijt}^m as a function of $(f_{it}^{m\prime}\Lambda_t^m, \eta_t^m)$, where η_t^m is a vector with elements $\{\eta_{jt}^m\}_{j\in\mathcal{J}^m}$ and Λ_t^m is a vector with elements $\{\lambda_{jt}^m\}_{j\in\mathcal{J}^m}$. We use $\{x_{jt}^m\}_{j\in\mathcal{J}^m}$ to refer to a vector containing the elements $(x_{1t}^m, \ldots, x_{Jt}^m)$ and we use $j \in \mathcal{J}^m$ to denote that the product index j is market-specific and hence it covers only the products in the set $\mathcal{J}^m$. Using the properties of the extreme value distribution we obtain

$$s_{ijt}^m(f_{it}^{m\prime}\Lambda_t^m, \eta_t^m) = \frac{\exp(f_{it}^{m\prime}\lambda_{jt}^m + \eta_{jt}^m)}{1 + \sum_{h\in\mathcal{J}^m}\exp(f_{it}^{m\prime}\lambda_{ht}^m + \eta_{ht}^m)}. \tag{6}$$

The overall market share, denoted by s_{jt}^m, of product j and time t in market m, measured over the entire population, is obtained by integrating $s_{ijt}^m(f_{it}^{m\prime}\Lambda_t^m, \eta_t^m)$ over the individual-specific parameters in f_{it}^m. Therefore we have that

$$s_{jt}^m = \int \frac{\exp(f_{it}^{m\prime}\lambda_{jt}^m + \eta_{jt}^m)}{1 + \sum_{h\in\mathcal{J}^m}\exp(f_{it}^{m\prime}\lambda_{ht}^m + \eta_{ht}^m)}\phi(f_{it}^m; \bar{f}_t^m, A^m\Psi A^{m\prime})df_{it}^m \tag{7}$$

If we use $f_{it}^m = \bar{f}_t^m + \upsilon_i^m$ and $\upsilon_i^m \sim \phi(0, A^m\Psi A^{m\prime})$ we can write equation (7) as

$$s_{jt}^m = \int \frac{\exp(\mu_{jt}^m + \lambda_{jt}^{m\prime}\upsilon_i^m)}{1 + \sum_{h\in\mathcal{J}^m}\exp(\mu_{ht}^m + \lambda_{ht}^{m\prime}\upsilon_i^m)}\phi(\upsilon_i^m; 0, A^m\Psi A^{m\prime})d\upsilon_i^m, \tag{8}$$

where $\mu_{jt}^m = (\bar{f}_t^m)'\lambda_{jt}^m + \eta_{jt}^m$. Note that the share s_{jt}^m inherits randomness only from the term η_{jt}^m as we integrate over υ_i^m.

Following Jiang et al. (2009) we denote the relationship between the shares

vector $s_t^m = \{s_{jt}^m\}_{j \in \mathcal{J}^m}$ and the vector with aggregate demand shocks η_t^m in (7) as

$$s_t^m = h(\eta_t^m | \Lambda_t^m, \bar{f}_t^m, \Psi). \tag{9}$$

Based on the relation in (9) and the distribution of η_t^m, the joint density of the shares at time t is

$$\pi(s_t^m | \Lambda_t^m, \bar{f}_t^m, \Psi, \tau_m^2) = \phi(h^{-1}(s_t^m | \Lambda_t^m, \bar{f}_t^m, \Psi) | 0, \tau_m^2) | J_{s_t^m \to \eta_t^m}|^{-1}, \tag{10}$$

for $t = 0, \ldots, T$ and $m = 1, \ldots, M$ and where we use $\pi(.)$ to denote a generic density and $\pi(y|x)$ the density of y given x. In addition, the Jacobian $J_{s_t^m \to \eta_t^m}$ is defined as the $(J^m \times J^m)$ matrix with elements

$$\partial s_{jt}^m / \partial \eta_{kt}^m = \begin{cases} -\int s_{ijt}^m s_{ikt}^m \phi(\upsilon_i^m; 0, A^m \Psi A_m') d\upsilon_i^m & \text{if } k \neq j \\ \int s_{ijt}^m (1 - s_{ikt}^m) \phi(\upsilon_i^m; 0, A^m \Psi A_m') d\upsilon_i^m & \text{if } k = j, \end{cases} \tag{11}$$

where the arguments of the functions s_{ijt}^m and s_{ikt}^m are dropped for convenience, see (8) and $j, k \in \mathcal{J}^m$. Given equation (10) the joint conditional density for the shares, or the likelihood, for market m is given by

$$\pi(s^m | \Lambda^m, \bar{f}^m, \Psi, \tau_m^2) = \prod_{t=1}^{T} \pi(s_t^m | \lambda_t^m, \bar{f}_t^m, \Psi, \tau_m^2), \tag{12}$$

where $\Lambda^m = (\Lambda_1^m, \ldots, \Lambda_T^m)$, $s^m = (s_1^m, \ldots, s_T^m)$ and $\bar{f}^m = (\bar{f}_1^m, \ldots, \bar{f}_T^m)$.

Two difficulties in this model are the inversion of the share function $h()$ in equation (9) and the evaluation of the integrals in equations (8) and (11). The inversion and the integration are required to obtain the aggregate shocks η_t^m and hence to evaluate the density in equation (10). We discuss these two issues next.

2.2 Share inversion method and integral approximation

To calculate the joint density in (12) we need to take two hurdles. First we need to solve the integrals in (8) and (11). Next, we need to obtain the inverse of the

function $h()$ in (9).

We apply the contraction mapping of Berry et al. (1995) to obtain the inverse in terms of μ_{jt}^m for all necessary m, j and t. Within this procedure we need to calculate the market shares given μ_{jt}^m, Λ_t^m and Ψ, $j \in \mathcal{J}^m$, $t = 1, \ldots, T$ and $m = 1, \ldots, M$ by integrating equation (8) with respect to υ_i. We numerically approximate this integral by averaging over H draws from the distribution of υ_i that is $N(0, A^m \Psi A^{m\prime})$. Jiang et al. (2009) report that H ranges from 20 to 50 in previous literature and they show that their Bayesian estimator has the same performance for $H = 50$ and $H = 200$. However, in our case we may need more draws as we develop the model for many more parameters.

A common approach to obtain each of the H draws of υ_i is based on the product of the Cholesky decomposition of $A^m \Psi A^{m\prime}$ and draws from a standard normal, that is $\upsilon_i^d = (A^m \Psi A^{m\prime})^{1/2} \zeta^d$ where $\Sigma^{1/2}$ denotes the Cholesky decomposition of Σ and $\zeta^d \sim N(0, \mathbf{I})$ for $d = 1, \ldots, H$, where $\mathbf{I}$ denotes an identity matrix. A more efficient approximation of the integral may be obtained by using a quasi-random scheme to generate the ζ^d. Train (2003, chap. 9, page 236) suggests scrambled Halton sequences for integrals of large dimensions and his suggestion, we believe, is motivated by the same family of logit models that we are concerned with here.

In Figure 1 we compare the integration results based on scrambled Halton sequences versus the integration results based on regular normal draws. We consider the scenario where the parameters are known and we use the approximation method discussed above to obtain the market shares. In the top panel we report the performance when the integral has only three dimensions and in the lower panel we report the performance when the integral has 30 dimensions. In both panels we report the market share for only one of the products. This simple exercise suggests that the market shares are much better approximated by integrating with Halton draws regardless of the dimension of the integral. For large dimensions the approximation of the normal draws seems to converge to the approximation of the Halton draws after the number of draws (H) is higher than 400 while the approximation based on Halton draws performs well for $H > 100$.

3 Bayesian Inference

In this section we discuss the priors we choose to complete the model specification. Specifically, we present in subsection 3.1 a prior for the matrix Ψ that is simple to calibrate when analyzing many products and markets and at the same time the prior will let us treat the scale and the correlation structure of Ψ separately. Next in subsection 3.2, we discuss the market-specific priors. Finally in subsection 3.3, we discuss the MCMC sampling scheme.

3.1 Prior and Structure for Ψ

Jiang et al. (2009) specify the covariance matrix Σ^m in terms of the unique elements of its Cholesky root. They set $\Sigma^m = U'U$ where

$$U = \begin{pmatrix} e^{r_{11}} & r_{12} & r_{13} & \dots & r_{1J} \\ 0 & e^{r_{22}} & r_{23} & \dots & r_{2J} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & e^{e_{J-1,J-1}} & r_{J-1,J} \\ 0 & 0 & \dots & 0 & e^{r_{JJ}} \end{pmatrix}, \tag{13}$$

and they choose to set separate normal priors for the diagonal and off-diagonal elements of U. That is, Jiang et al. (2009) set $r_{jj} \sim N(0, \sigma_j^2)$ for the diagonal elements $j = 1, \dots, J$ and $r_{jk} \sim N(0, \sigma_{od}^2)$ for the off-diagonal elements $j \neq k$. Note that Jiang et al. (2009) deal with only one market m and that in our approach we model the heterogeneity through Ψ (that is a $K \times K$ matrix) and not through Σ^m (that is a $J \times J$ matrix) where K is the number of factors while J is the number of products.

This prior specification enforces the positive-definitiveness of Σ^m and in addition the priors on the r_{jk} elements used by Jiang et al. (2009) are symmetric and this matched well with the random walk Metropolis Hastings [MH] sampling scheme they propose to sample the r-parameters. A second advantage of Jiang et al. (2009) prior is that it can be set to be relatively uniform on the correlation range

$(-1, 1)$. Overall, this specification is attractive and simple but it also has a number of shortcomings. First, Jiang et al. (2009) note that to obtain a plausible (implied) prior on the variances in Σ^m the prior variances σ_j^2 should be decreasing with j and they provide a particular relation between σ_j^2 and j. However, the prior in one of the elements of (13) may affect many of the elements in Σ^m and this complicates the prior interpretation, specially when J is large. Second, this prior imposes a correlation structure simultaneously with the overall scale of the heterogeneity. Other studies point out that it may be relatively hard to identify the heterogeneity (Bodapati and Gupta (2004)) and the uncertainty related to the Σ^m elements is therefore usually large, see for example Jiang et al. (2009) and Musalem et al. (2006). However, we do not know if the large uncertainty reported in previous studies is due to the uncertainty on the overall scale of Σ^m or if it is due to the correlation structure in Σ^m. Finally, it is well known that the correlation structure of Σ^m is very important in order to obtain different substitution patterns far from the IIA assumption of the logit. Therefore, we would like to use a prior that can let us deal with the scale and correlations separately.

Finally, changing one element of U may lead to a very different Σ^m. This fact makes the implementation of an efficient MH sampler difficult if J is large. More precisely, in their MH scheme Jiang et al. (2009) choose to draw the candidate elements of U from a multivariate normal that is calibrated based on a short chain of their model MCMC output. The length of the chain needed for the calibration of the multivariate normal needs to be longer when the number of dimensions is large. When dealing with large dimensions, the step size in the random walk MH sampling, for each of the r_{jk} elements, needs to be set smaller as the dimension increases in order to keep a good acceptance rate in the estimation algorithm. Although this last issue always arises whenever many elements are updated simultaneously, here it can be more dramatic as "local" changes in U lead to "global" changes in Σ^m.

Summarizing, we would like to use a prior specification that is simpler to calibrate when dealing with large dimensions and at the same time we like to treat the scale and the correlation structure of the heterogeneity separately.

We choose to use the prior specification of Barnard et al. (2000) for Ψ. We define $\Psi = DRD$ where D is a diagonal matrix with K elements (standard deviations) and R is a $K \times K$ correlation matrix. For the variances in D we set the prior $\log(\text{diag}(D)) \sim N(0, \Delta)$. Formulating a prior on R is not straightforward because we need a prior that deals with all the elements of R and the restrictions on them. We need to assure the positive definitiveness of R, the range of its elements must be $(-1, 1)$ and all the elements together should satisfy triangularity restriction inherent to any correlation matrix. In addition, we need to update all the elements of R simultaneously to ease the computational burden. However, based on any variance matrix Σ one can obtain the corresponding correlation matrix by standardization. Hence, we assume $R = f_c(S)$ and we specify an Inverted Wishart prior for S with parameters (G, v). The function $f_c()$ transforms a covariance matrix to a correlation matrix. The location matrix G is set such that the expected value of S is an identity matrix; that is $G = (v - h - 1) \times \mathbf{I}$, where h is the number of columns of R, v is the degrees of freedom of the Inverted Wishart and $\mathbf{I}$ is an identity matrix of size h. Note that our variance matrix Ψ is now actually a function of D and S. In the MCMC sampling below we will actually sample these two matrices.

Barnard et al. (2000) set a prior directly on R while we set a prior on R implicitly by the prior on S. The main reason why we deviate from them is that evaluation of the posterior is very costly in our application and hence we need to use a proposal for S that updates all the correlations in R in a single step. In contrast, the computation time of the application in Barnard et al. (2000) allows for a relatively fast element by element update of the matrix R.

In Figure 2 we report the implied correlation distributions for two different degrees of freedom and for two elements of a Ψ matrix of size 10×10. The implied correlations can be set to be roughly uniform on the $(-1, 1)$ interval depending on the degrees of freedom set on the Wishart distribution. Therefore, the implied correlations of this prior are very similar to the implied correlations of the specification used by Jiang et al. (2009). However, the parameters in our suggested priors are easier to interpret.

3.2 Market-Specific Priors and Joint Posterior

We presented the likelihood for each of the m markets in equation (12) and we presented the structure and prior for Ψ in the previous section. This variance matrix applies to all markets. What is left to specify are the priors for τ_m^2, Ω^m and the initial state distributions for all common dynamic factors f_t^m. In addition the matrix Γ_t^m needs to be defined. We define $f_1^m \sim N(0, P^m)$ where we set P^m reasonably large and non-informative, $P^m = 100\mathbf{I}$ for all m. We assume $\Omega^m = \sigma_m^2 \mathbf{H}_m$ where $\mathbf{H}_m$ is a diagonal matrix of size J^m and $\sigma_m^2 \sim v_o s_o^2 / \chi_{v_o}^2$. The diagonal elements of $\mathbf{H}_m$ are equal to one for the factors f_t^m that are time-varying and equal to zero for the factors that are fixed over time. We set Γ_t^m equal to an identity matrix $\mathbf{I}_m$ of size J^m and for τ_m^2 we do not set any prior.

The joint posterior is proportional to the product of the likelihood and priors for each market times the prior distribution of Ψ that apply to all markets. Note that the factor loadings are assumed to be given, as they represent observed product characteristics. The posterior becomes

$$\pi(\bar{f}^*, \tau_*^2, \sigma_*^2, D, S | s^*, \Lambda) \propto \pi(\log(\text{diag}(D)); 0, \Delta)\pi(S; \mathbf{I}, v) \times$$
$$\left(\prod_m \pi(s^m | \bar{f}^m, \Psi, \tau_m^2) \pi_m(\bar{f}_1^m; 0, P^m) \left[\prod_{t=1}^{T-1} \pi(\bar{f}_{t+1}^m | \bar{f}_t^m, \sigma_m^2 \mathbf{I}_m) \right] \pi(\sigma_m^2; v_o, s_o^2) \right), \quad (14)$$

where $s^* = (s_1, \ldots, s_M)$, $\bar{f}^* = (\bar{f}^1, \ldots, \bar{f}^M)$, $\tau_*^2 = (\tau_1^2, \ldots, \tau_M^2)$, $\sigma_*^2 = (\sigma_1^2, \ldots, \sigma_M^2)$, $\Psi = D f_c(S) D$.

In addition, the priors for f_1^m, σ_m^2, D and S are defined as follows

$$\begin{aligned}
&\pi(\bar{f}_1^m; 0, P^m) \sim N(\bar{f}_1^m; 0, P^m) \\
&\pi(\sigma_m^2; v_o, s_o^2) = (\sigma^2)^{-(v_o/2+1)} e^{-v_o s_o^2 / 2\sigma^2} \\
&\log(\text{diag}(D)) \sim N(0, \Delta) \\
&\pi(S; \mathbf{I}, v) \propto \frac{|\mathbf{I}|^{v/2}}{|S|^{(v+K+1)/2}} e^{-1/2 tr(S^{-1}\mathbf{I})}
\end{aligned} \quad . \quad (15)$$

3.3 MCMC algorithm

The approach we follow is a combination of the sampler proposed in Jiang et al. (2009) with simulation smoother of Durbin and Koopman (2002) and a Metropolis Hastings sampler for Ψ. We use the following steps: (i) conditional on Ψ we use the contraction mapping to obtain the (implied) μ_{jt}^m, for $m = 1, \ldots, M$, $j \in \mathcal{J}^m$, $t = 1, \ldots, T$; (ii) conditional on Ψ (and μ_{jt}^m) we use the simulation smoother to sample $\bar{f}^*$, the μ_{jt}^m values appear as dependent variables in this smoother; (iii) conditional on $\bar{f}^*$ and all μ_{jt}^m we sample τ_*^2 and σ_*^2; (iv) finally we use a Metropolis Hastings sampler to draw the elements of D and S which determine Ψ.

More specifically, we use the following three set of conditionals

$$\begin{aligned} &\bar{f}^*|\Psi, \sigma_*^2, \tau_*^2, s^*, \Lambda \\ &\sigma_*^2, \tau_*^2|\Psi, \bar{f}^*, s^*, \Lambda \\ &D, S|\sigma_*^2, \tau_*^2, \bar{f}^*, s^*, \Lambda. \end{aligned} \tag{16}$$

We draw the first set of conditionals using the simulation smoother of Durbin and Koopman (2002). That is, given μ_{jt}^m for all m, j and t we can draw the parameters of the following measurement and state equations

$$\begin{aligned} \mu_t^m &= \Lambda_t^{m\prime} \bar{f}_t^m + \eta_t^m \qquad \text{with} \quad \eta_{jt}^m \sim N(0, \tau_m^2) \\ \bar{f}_{t+1}^m &= \Gamma_t^m \bar{f}_t^m + \Pi_t^m \omega_t^m \quad \text{with} \quad \omega_t^m \sim N(0, \sigma_m^2 \mathbf{I}_m), \end{aligned} \tag{17}$$

where μ_t^m is defined as the $(J_m \times 1)$ vector with elements μ_{jt}^m, $j \in \mathcal{J}^m$. This specification is attractive because we can set some of the common factors $\bar{f}_t^m$ to be fixed in time while others can remain time-varying. This is done simply by setting some of the elements in the diagonal matrix Π_t^m equal to zero. The simulation smoother of Durbin and Koopman (2002) gives a draw from the joint posterior of $\bar{f}_t^m$, for $t = 1, \ldots, T$. For details we refer to their paper. Conditional on μ_t^m and $\bar{f}^m$ sampling the variances is straightforward. Given our priors they can be sampled from Inverted Gamma distributions. That is, $\tau_m^2 \sim IG(n_\tau^m, s_\tau^2)$ and $\sigma_m \sim IG((\upsilon_o + n_\sigma^m), (s_\sigma^2 + s_o^2)$. The n_τ^m are the number of observations available for the

measurement equation at market m for $m = 1, \ldots, M$ and s_τ^2 is the sum of squared residuals of the measurement equation. The n_σ^m is the number of observations available in the state equation at market m and s_σ^2 is the sum of squared residuals of the measurement equation. The v_o and s_o^2 are the parameters of the prior for the variance of the state equation, see the priors in equation (15).

For the third set of conditionals we use a Metropolis Hastings algorithm. For the (log of the) elements of D we use a standard random walk as candidate distribution. For comparison with the second part of this step we write the proposal as

$$\log(\text{diag}(D^{\text{candidate}})) \sim N(\log(\text{diag}(D^{\text{current}})), \zeta^2 \mathbf{I}). \tag{18}$$

For S we also propose a random walk candidate distribution. However, for efficiency in the total sampler we which to have a candidate that can generate matrices close to the current value. As a candidate distribution we use an inverted Wishart distribution which has the current value as expected value, that is,

$$S^{\text{candidate}} \sim IW((v_1 - K - 1)S^{\text{current}}, v_1). \tag{19}$$

We choose v_1 and ζ^2 to achieve between 20% and 50% acceptance rate in the Metropolis steps. In the MCMC we use two Metropolis steps to update D and S separately.

To sample D and S we evaluate the model posterior in equation (14) in two Metropolis steps, the first for D and the second for S. We set

$$D^{new} = D^{cand} \quad \text{with probability} \quad \min\{\tfrac{p^*(D^{cand}|S,\bar{f}^*,\tau_*^2,\sigma_*^2,s^*,\Lambda)}{p^*(D^{prev}|S,\bar{f}^*,\tau_*^2,\sigma_*^2,s^*,\Lambda)}, 1\}, \tag{20}$$

and we set

$$S^{new} = S^{cand} \quad \text{with probability} \quad \min\{\tfrac{p^*(S^{cand}|D,\bar{f}^*,\tau_*^2,\sigma_*^2,s^*,\Lambda)}{p^*(S^{prev}|D,\bar{f}^*,\tau_*^2,\sigma_*^2,s^*,\Lambda)}, 1\}. \tag{21}$$

The candidate and previous posterior density in equation (20) are given by

$$p^*(D^{cand}|S, \bar{f}^*, \tau_*^2, \sigma_*^2, s^*, \Lambda) = \\ \pi(\log(\text{diag}(D^{cand})); 0, \Delta) \times \prod_m \pi(s^m|\Lambda^m, \bar{f}^m, \Psi^{cand}, \tau_m^2), \quad (22)$$

and by

$$p^*(D^{prev}|S, \bar{f}^*, \tau_*^2, \sigma_*^2, s^*, \Lambda) = \\ \pi(\log(\text{diag}(D^{prev})); 0, \Delta) \times \prod_m \pi(s^m|\Lambda^m, \bar{f}^m, \Psi^{prev}, \tau_m^2). \quad (23)$$

where $\Psi^{cand} = D^{cand} f_c(S) D^{cand}$ and $\Psi^{prev} = D^{prev} f_c(S) D^{prev}$ while the terms in the Metropolis step in equation (21) are given by

$$p^*(S^{cand}|D, \bar{f}^*, \tau_*^2, \sigma_*^2, s^*, \Lambda) = \\ \pi(S^{cand}; \mathbf{I}, v) \times \pi(S^{prev}; S^{cand}, v) \times \prod_m \pi(s^m|\Lambda^m, \bar{f}^m, \Psi^{cand}, \tau_m^2), \quad (24)$$

and by

$$p^*(S^{prev}|D, \bar{f}^*, \tau_*^2, \sigma_*^2, s^*, \Lambda) = \\ \pi(S^{prev}; \mathbf{I}, v) \times \pi(S^{cand}; S^{prev}, v) \times \prod_m \pi(s^m|\Lambda^m, \bar{f}^m, \Psi^{prev}, \tau_m^2), \quad (25)$$

where $\Psi^{cand} = D f_c(S^{cand}) D$, $\Psi^{prev} = D f_c(S^{prev}) D$ and $\pi(s^m|\Lambda^m, \bar{f}^m, \Psi, \tau_m^2)$ is defined in equation (12). We use the proposal distributions in equation (18) and equation (19) to draw the candidate matrices D^{cand} and S^{cand} based on their previous values D^{prev} and S^{prev}, respectively.

The Metropolis steps are very costly in terms of computation time in the MCMC algorithm. This is the only step in the algorithm where we need to use the BLP contraction mapping and where we need to evaluate the Jacobian in equations (10) and (11). Some time may be saved by jointly updating these matrices. However, the joint updating of D and S will not let us distinguish what is driving the acceptance

rate in the Metropolis steps. Moreover, the separate updating of these matrices let us distinguish if a candidate matrix Ψ^{cand} ($\Psi = Df_c(S)D$) is rejected because of its correlation structure or because of its overall scale.

4 Simulation Experiment

We test our modeling approach on simulated data and in this section we discuss the data generation process and the results of the MCMC estimation procedure.

4.1 Data Simulation

In this section we describe how we create synthetic data and we consider a setting where we have data for many products and markets. This setting is not typical in the literature but it that corresponds with the setting that we deal with in the application.

We assume products are sold in 10 markets and we simulate 4 years of monthly data for each market. Each market will be assigned a specific number of products and these products will be assigned to 10 different brands.

All 10 brands are available in each market and we assign 5, 6, 8 or 10 products to each brand at each market. Hence, the number of products assigned to a brand varies per market and each market consists of a specific number of products. The probability of a brand to be assigned 5 or 6 products at each specific market is 90% while the probability of being assigned 8 or 10 products is 10%. That is, the expected number of products per brand is 5.85 and the expected number of products per market is 58.5. This is a large number of products relative to previous studies. For example, Jiang et al. (2009) and Yang et al. (2003) study one market that consists of 3 products and one outside good while Musalem et al. (2006) apply their model to a setting with four products and one outside good.

The mean utilities (the μ_{jt}^m) for products are market-specific. The mean products' utilities are assumed to depend on the products' brands, the products' attributes and the products' prices and promotions. We define 5 attributes and we

assign only one attribute per product and all attributes are equally likely per product. Note that this last assumption implies that each brand may have a certain number of products that share the same attributes. We set one of the attribute coefficients as the base and equal to 0 while the rest is generated from a normal distribution with mean 0 and variance 5.

Next, we define the price and promotion coefficients and these are -5 and -2, respectively, and these coefficients are the same across markets. The price series for each product follows a sine curve (with a very long cycle) plus normal noise with mean 2 and variance 1. To create the promotional series we use a uniform distribution with range $(0, 1)$. We assume that there is a 30% chance of a promotion and the range of promotions is between 0 and 30%. That is, when we draw a promotional index value (from the uniform) lower than 0.70 then the promotional index is equal to 1 otherwise the promotion index is equal to the drawn value.

Attribute, price and promotions coefficients will be fixed in time while the brand coefficients will be time-varying. We generate 10 brand coefficients using the recursion in equation (17) and we set σ_m to be equal to 0.40. We set the initial values for the brand coefficients f_1^m based on a normal distribution with mean -3 and variance 0.16. We use the same recursion to generate the attribute, price and promotion coefficients and their initial value is assigned as we discussed in the previous paragraph. We further need to set Π_t^m to be a diagonal matrix with the first 10 elements equal to 1 and the remainder 6 elements of the diagonal are equal to 0. The Γ_t^m is equal to an identity matrix of size 16.

The factor loadings Λ_t^m will consist of brand and attribute dummies for all products at time t plus the products prices and promotions at time t. That is, Λ_t^m is a $J \times K$ matrix, J is the number of products and K is equal to 16 (the number of brands, attribute, price and promotion coefficients). Finally, we assume that the variance of product demand shocks τ_m^2 are equal to 0.8 for all m.

We use the specification of $\Psi = DRD$ to draw the random coefficients υ_i^m. We first draw a matrix P based on a $IW(I_{16}, 21)$ and we set $R = f_c(P)$. The implied range of the correlations in R goes from -1 to 1 but the extremes of

the range are not common. Further, we assume the scale of the heterogeneity depends on both small and large elements with the purpose of checking whether their size affects their retrieval from the synthetic data. That is, we set $D^2 = (2, 2, 2, 2, 8, 8, 8, 8, 4, 4, 4, 4, 2, 2, 2, 2)$. Finally, we use 3000 draws to approximate the integral in equation (8) and we generate the draws of the random coefficients based on the Cholesky decomposition of Ψ and normal draws generated with scrambled Halton sequences. The 16 factors in Ψ are available at every market and therefore the A^m matrix is the same for all markets and it is equal to I_{16}.

4.2 MCMC Setup

We use a hybrid Metropolis Gibbs sampler to estimate the parameters of the model in equation (7). The sampler iterates over the conditionals in equation (16). The first set of conditionals concerns the $\bar{f}^*$. We set the prior on the initial values as $f_1^m \sim N(0, 100\mathbf{I})$ for all m and we use the simulation smoother of Durbin and Koopman (2002) to sample all elements of $\bar{f}^*$.

The second set of conditionals samples the variances of equation (17). We did not set any prior information on σ_m^2 and τ_m^2 for all m. Hence, $\tau_m^2 \sim IG(n_m, s_m)$ where n_m are the number of observations in the measurement equation in (17) and s_m are the sum of squared residuals in the same equation. In a similar fashion, $\sigma_m^2 \sim IG(n_m^s, s_m^s)$ where n_m^s are the number of observations and s_m^s are the sum of squared residuals of the state equation in (17).

The third set of conditionals concerns the sampling of the D and S matrices. We set the v parameter in the prior $\pi(S; \mathbf{I}, v)$ equal to 21 and we use $\Delta = 10\mathbf{I}$ in the normal prior of the log of the diagonal elements of D.

We use the proposal distributions in equation (18) and equation (19) to draw the candidate matrices D^{cand} and S^{cand}, respectively. In these proposals we set ζ^2 equal to 0.01 and υ_1 equal to 10000. The large number in υ_1 corresponds to steps of approximately 0.05 in the elements of the correlation matrix R where $R = f_c(S)$. We calibrated ζ^2 and υ_1 to achieve an acceptance rate between 20% and 50% for

both Metropolis steps.

We let the Gibbs-Metropolis sampler to run for 20 thousand iterations. However, we do oversampling of Ψ. We use 4 updates of S and one of D at every iteration. That is, we generate 100 thousand candidates for the matrix Ψ. The matrix Ψ contains 136 unique elements and our purpose with the oversampling is to let all these elements to move at larger steps at every iteration and let them adjust better to the rest of the model parameters drawn at every iteration. In this way, the oversampling may compensate for the small moving steps that we need to achieve a good acceptance rate in the Metropolis algorithm. Haran et al. (2003) also consider the oversampling of parameters to accelerate their computation in the MCMC algorithm.

4.3 Results of the Simulation Experiment

In Table 1 we present the posterior mean and the 99% Highest Posterior Density Region (HPDR) of the demand shocks for every market. The true value of τ_m^2 is equal to 0.66 for all markets. Note that we generated data for 10 markets. In most cases, the posterior mean is very close to its true value. The maximum absolute deviation of the posterior mean from the true value is approximately 0.06, see the $\tau_{m=6}^2$ that is equal to 0.580.

In Table 2 we present the posterior mean and HPDR of the variance term in the state equation (17), that is σ_m^2. The true value of this parameter is 0.16 while in most cases the posterior mean is close to 0.12. That is, we are finding a small negative bias that is close to 0.04 for most cases.

In Figure 3 we present the estimates of the fixed coefficients (in circles) and the box-plots of their posterior distribution. Note that we specified 4 attribute coefficients and one price and promotion coefficient that vary across markets. That is a total of 50 coefficients in all markets. We see that for 30 out of the 50 coefficients the circles (true values) overlap with the position of their distribution in the box-plot. In the same figure, we see that there is a systematic positive bias in the

posterior distribution of the price coefficients. The true value of the price coefficient is equal to -5 while the posterior distribution is higher than -5. In contrast, the posterior distribution of the promotion coefficients overlap with its true value (-2) for all markets.

In Figure 4 we present the distribution of the time-varying brand coefficients for the 5^{th} market. We see that the overall time profile is well retrieved by the estimation algorithm. In most cases the true value is inside the 99% HPDR. The results for the other markets are very similar.

In Table 3 we report the posterior mean and HPDR for the elements of the D^2 matrix. We see that the 99% HPDR contains the true value for 7 out of the 16 elements. The deviation of the posterior mean from its true value, when the true value is not contained in the HPDR, may be as small as 1 or as large as 6 variance points. That is, we find large uncertainty regarding the scale of the heterogeneity driven by the random coefficients. Jiang et al. (2009), Musalem et al. (2006) and Yang et al. (2003) find similar levels of uncertainty.

In Figure 5 we report the 99% HPDR (in dashed lines) and the true value (solid line) of the 120 unique elements of the correlation matrix R ($f_c(S)$). We find that the HPDR contains the true value for 57 out of 120 elements, that is 47.5% of the elements. However, we find that the posterior mean of the correlations is on average 0.16 points far from its true level. Hence, our results suggest that the uncertainty regarding the scale of the heterogeneity (the elements of D) is much larger than the uncertainty in the elements of the correlation matrix R.

5 Empirical Application

In this section we apply our estimation approach to a real dataset and we analyze the substitution patterns between a large number of products. Next we provide a description of the data (in subsection 5.1), the modeling details (in subsection 5.1) and the estimation results (in subsection 5.3).

5.1 Data

Our dataset contains sales, price and promotion data for all the products of one supermarket food category. The data is monthly and it covers a period of four years and 18 different regions. Consumers at each region may have available a minimum of 25 up to a maximum of 65 products of 20 different brands. Each brand has its own positioning in terms of calories, taste and labeling while each brand may offer products of the same size and packaging. Therefore, we can describe each product in terms of its brand, size and packaging attributes and its price and promotion data. There are brands with similar attributes both in terms of calories and taste and in terms of packaging and size and these brands are usually produced by different companies. Our data contains products sold by all major companies at each market and very few firms compose the market.[1] Depending on the market, the size of the outside good varies from 20 up to a maximum of 50%. The calculation of the outside good share is region-specific and it varies according to the share of the closest and competing food categories.

5.2 Modeling Details and MCMC Setup

The MCMC setup for the application is very similar to the MCMC setup we use for the simulated data. An important distinguishing feature is that the matrix Ψ consists of 32 rows and columns. This number corresponds to 20 brands, 11 size and packaging attributes, one price and one promotion factor. We leave one attribute as reference and this results in 32 random coefficients. In the application the A^m matrices select the appropriate elements of the Ψ matrix relevant for the market m. That is, some attributes or brands may not be available in all markets.

We will assume that all coefficients are fixed with the exception of the brand coefficients that will be specified as time-varying. We use the priors in equation (15) where we set $P^m = 100\mathbf{I}$ for all m, $v_o = 1$ and $s_o = 0.01$. The Δ matrix is equal to $25\mathbf{I}$ and $v = 35$. We did not set a prior on τ_m^2 parameters. The proposal

[1] Because of our confidentiality agreement we can not reveal the companies names, brands or any other product or market information in the paper.

distributions in equation (18) and equation (19) have the parameters $v_1 = 30000$ and $\zeta^2 = 1/200$. This configuration achieves between 30% and 50% of acceptance rate in the Metropolis updates of S and D. We sample D and S separately in the same way as we did in the simulation experiment.

The matrix Π_t^m in equation (17) is set equal to an identity matrix of size K^m (K^m is the number of factors at each m) and we set some of its diagonal elements equal to 0 and these zeros correspond to the factors related to size, packaging and to price and promotions. The matrix Γ_t^m is of size J^m (the number of products available at market m) times K^m and it is also set to be an identity matrix.

We ran the MCMC chain for 50 thousand iterations and we discarded the first 10 thousand with a thin value of 20. The computation time was of approximately five days. The number of draws that we used for approximating integrals was 200 and we use draws based on scrambled Halton sequences.

5.3 Estimation Results

We present the posterior mean and HPDR of the τ_m^2 parameters in Table 4. The uncertainty of the demand shocks is very large for six of the eighteen markets, see the τ_m^2 for m equal to 1, 2, 12, 13 and 17. The uncertainty in the demand shock for the remaining markets seems small relative to these six markets.

The posterior mean of HPDR of the σ_m^2 parameters can be read in Table 5. These are the variances of the time-varying coefficients and we see that they are very small as we expected. The variance of time-varying parameters in state space models is usually small (Fruhwirth-Schnatter, 2004) and this indicates slowly evolving factors.

In Table 6 we present the posterior mean and HPDR for the fixed coefficients at three markets. We notice that price and promotion coefficients have the expected negative signs. The promotional index is a number that takes a value between 0 and 1 and it indicates the percentage of the regular price level that is observed. We notice that the uncertainty related to the price coefficients varies across markets

while at the same time they remain negative. The preference for *size* also varies per market and we find that for each market there are only two *sizes* with a positive posterior mean that may be larger than the base category. In Figure 6 we report the evolution of the time-varying brand coefficients. We report the time profiles of the time-varying factors relative to their starting point and their corresponding 99% HPDR. This transformation is useful to illustrate how some brands' preferences (measured by the time-varying factors) face large variations relative to their starting position, like brands C, E, F or L, while we see other brands like J or B with much smaller time variation. Note that this figure does not show the level uncertainty around the time-varying brand coefficients. Their level uncertainty, however, is similar to the uncertainty of the fixed coefficients. We see also that the coefficients for different types of packaging show significant time variation relative to their starting point, see the bottom row in Figure 6.

In Figure 7 we report the distribution of 60 elements of the Ψ matrix. The Ψ matrix size is 32×32 and therefore it contains 528 unique elements. We notice that the uncertainty varies per element but overall the uncertainty is relative small for most correlations. The element 39 in the lower panel has the largest uncertainty and its range goes from -0.75 up to 0 while there are other cases like the element 12 in the upper panel with very tight posterior distributions.

In Table 7 we present the posterior mean and the 99% HPDR of the matrix D^2. Some of the elements of the matrix are retrieved with a lot of uncertainty. For example, the posterior mean of the D^2_9 is 4.64 but its HPDR includes values close to 10 while the posterior mean of D^2_{15} is equal to 10.625 and its HPDR includes values as high as 22. These rest of the elements in the D^2 matrix, and the mayority, show a much smaller uncertainty relative to these high values in D^2_9 and D^2_{15}. Previous studies, like Jiang et al. (2009), Musalem et al. (2006) and Yang et al. (2003), report simular range of both the scale of the heterogeneity and its uncertainty.

Finally, in Figure 8 and Figure 9 we present the own-price and cross price elasticities for the products in market 2. We computed the elasticities as we describe in the Appendix A. The price elasticities have a range that goes from -1 to -3.5

while the cross-price elasticities range goes from 0 up to 1.6. In this last figure light (white) colors represent high values while darker (dark red) colors represent lower cross price elasticities. In Figure 9 we notice that many products respond to the price changes of a relatively small set of products. For example, a price change in the 10^{th} product affects almost all products in this market and their cross price elasticity is close to 1.66. Finally, we notice that substitution patterns (measured by cross price elasticities) are stronger among a small subset of products.

6 Conclusions

The estimation of aggregate share models based on the random coefficient logit specification presents different challenges. The scalability of models and estimation algorithms is one of these main challenges. Berry and Pakes (2007) is a recent paper with a similar concern as ours and that is the practical application of this family of models to larger and more comprehensive datasets. In this paper we investigate the scalability of the BBLP approach and we successfully applied our method to simulated data and to a relative large real dataset. It is large in terms of the number of products, brands and markets that it includes while it is still small relative to the time periods we have available.

Our specification is based on the recent advances of Jiang et al. (2009), Durbin and Koopman (2002) and Barnard et al. (2000). These advances all put together allow us to model time variation in preferences and to separate the uncertainty of the random coefficients in terms of their scale and in terms of their correlation. In addition, our mode specification combines global and market specific priors and this allows us pool information across several markets.

We believe that the uncertainty related to the random coefficients is a great challenge. In contrast with previous studies we report the uncertainty related to the correlation and the scale of the random coefficients separately. Our results point that the overall scale of the covariance matrix of the random coefficients may present a larger uncertainty relative to the uncertainty present in their correlation

structure. This last result is an initial step towards the untangling and modeling of the sources of uncertainty in the random coefficients of the BBLP approach and we consider that this is a promising area for further research.

We present an approach that is the "augmented" version of the BBLP and it should be considered whenever there is a large dataset of market shares available for analysis. Large datasets, particularly of shares, are rarely collected but they are becoming increasingly common and more detailed. Therefore, approaches like ours may be needed more often in the future.

We presented our results to managers and they showed a great interest in understanding the uncertainty regarding the correlations between a reduced number of key product factors. Their immediate questions concerned what factors are "competing" between each other and to what extent. Moreover, their intuition and knowledge of the market supports the idea that preferences for key factors, like brands, are evolving in time. However, they usually measure these time variations based on market wide "top of mind" surveys while the use of sales data for this type of analysis is rare. Hence, the modeling of the evolution in brands-preferences based in market shares data, they argue, is one of the key and most valuable aspects of our approach.

7 Tables and Figures

	Posterior	HPDR	
	Mean	1%	99%
$\tau^2_{m=1}$	0.636	0.618	0.698
$\tau^2_{m=2}$	0.640	0.624	0.670
$\tau^2_{m=3}$	0.595	0.580	0.626
$\tau^2_{m=4}$	0.659	0.641	0.717
$\tau^2_{m=5}$	0.618	0.598	0.657
$\tau^2_{m=6}$	0.580	0.565	0.598
$\tau^2_{m=7}$	0.657	0.640	0.711
$\tau^2_{m=8}$	0.662	0.647	0.755
$\tau^2_{m=9}$	0.641	0.624	0.682
$\tau^2_{m=10}$	0.630	0.615	0.680
Notes: The true value of τ^2_m is equal to 0.64 for all m. HPDR stands for Highest Posterior Density Region.			

Table 1: Simulation Experiment: Posterior Distribution of the Variance of the Demand Shocks

	Posterior	HPDR	
	Mean	1%	99%
$\sigma^2_{m=1}$	0.107	0.082	0.148
$\sigma^2_{m=2}$	0.120	0.096	0.157
$\sigma^2_{m=3}$	0.118	0.088	0.144
$\sigma^2_{m=4}$	0.120	0.094	0.155
$\sigma^2_{m=5}$	0.136	0.104	0.182
$\sigma^2_{m=6}$	0.134	0.102	0.170
$\sigma^2_{m=7}$	0.128	0.096	0.170
$\sigma^2_{m=8}$	0.116	0.090	0.153
$\sigma^2_{m=9}$	0.121	0.090	0.160
$\sigma^2_{m=10}$	0.128	0.098	0.170

Notes: The true value of σ^2_m is equal to 0.16 for all m. HPDR stands for Highest Posterior Density Region.

Table 2: Simulation Experiment: Posterior Distribution of σ^2_m

	Posterior	HPDR		Real
	Mean	1%	99%	Value
Brand A	1.950*	1.081	4.089	2.0
Brand B	0.435*	0.209	3.446	2.0
Brand C	3.036*	1.958	3.755	2.0
Brand D	1.127*	0.890	3.514	2.0
Brand E	7.890*	5.126	8.983	8.0
Brand F	3.929	3.119	4.702	8.0
Brand G	3.066	2.421	4.500	8.0
Brand H	1.992	1.776	3.789	8.0
Brand I	2.295	1.817	3.617	4.0
Brand J	2.219	1.938	3.603	4.0
Attribute b	4.827	4.270	5.316	4.0
Attribute c	1.990	1.463	3.636	4.0
Attribute d	1.473*	0.732	3.202	2.0
Attribute e	2.988	2.693	3.490	2.0
Price	0.866	0.543	1.510	2.0
Promotion	1.533*	1.112	2.550	2.0

Note: * means that the real value is included in the HPDR. HPDR stands for Highest Posterior Density Region.

Table 3: Simulation Experiment: Posterior Distribution of the elements of D^2.

	Posterior	HPDR	
	Mean	1%	99%
$\tau^2_{m=1}$	5.329	0.606	53.233
$\tau^2_{m=2}$	7.518	4.782	11.322
$\tau^2_{m=3}$	0.814	0.659	1.034
$\tau^2_{m=4}$	0.895	0.684	1.172
$\tau^2_{m=5}$	1.500	1.127	2.042
$\tau^2_{m=6}$	1.284	0.989	1.715
$\tau^2_{m=7}$	0.831	0.577	1.281
$\tau^2_{m=8}$	0.638	0.574	0.734
$\tau^2_{m=9}$	0.838	0.754	1.013
$\tau^2_{m=10}$	0.467	0.395	0.589
$\tau^2_{m=11}$	1.015	0.805	1.322
$\tau^2_{m=12}$	5.664	1.632	65.241
$\tau^2_{m=13}$	2.631	1.906	3.652
$\tau^2_{m=14}$	0.752	0.672	0.901
$\tau^2_{m=15}$	1.917	1.490	2.572
$\tau^2_{m=16}$	1.439	1.316	1.607
$\tau^2_{m=17}$	4.144	0.602	44.067
$\tau^2_{m=18}$	1.612	1.328	2.101
Note: HPDR stands for Highest Posterior Density Region.			

Table 4: Application: Posterior Mean and HPDR of the τ^2_m.

	Posterior	HPDR	
	Mean	1%	99%
$\sigma^2_{m=1}$	0.0149	0.0090	0.0249
$\sigma^2_{m=2}$	0.0200	0.0102	0.0397
$\sigma^2_{m=3}$	0.0628	0.0357	0.1213
$\sigma^2_{m=4}$	0.0384	0.0206	0.0650
$\sigma^2_{m=5}$	0.0128	0.0080	0.0276
$\sigma^2_{m=6}$	0.0180	0.0101	0.0293
$\sigma^2_{m=7}$	0.0361	0.0208	0.0600
$\sigma^2_{m=8}$	0.0322	0.0190	0.0563
$\sigma^2_{m=9}$	0.0390	0.0202	0.0654
$\sigma^2_{m=10}$	0.0339	0.0196	0.0611
$\sigma^2_{m=11}$	0.0557	0.0272	0.1021
$\sigma^2_{m=12}$	0.0131	0.0072	0.0236
$\sigma^2_{m=13}$	0.0204	0.0113	0.0425
$\sigma^2_{m=14}$	0.0252	0.0150	0.0450
$\sigma^2_{m=15}$	0.0132	0.0088	0.0230
$\sigma^2_{m=16}$	0.0255	0.0139	0.0429
$\sigma^2_{m=17}$	0.0123	0.0073	0.0249
$\sigma^2_{m=18}$	0.0147	0.0087	0.0202
Note: HPDR stands for Highest Posterior Density Region.			

Table 5: Application: Posterior Mean and HPDR of the σ^2_m.

		Posterior	HPDR	
		Mean	1%	99%
Market 1	Size A	0.811	-0.098	1.766
	Size B	-1.835	-3.025	-0.577
	Size C	0.268	-0.643	1.164
	Size D	-0.513	-2.131	1.174
	Size E	-2.042	-3.053	-1.017
	Price	-2.684	-3.925	-1.529
	Promotion	-2.380	-6.455	1.856
Market 2	Size A	0.120	-0.207	0.438
	Size B	-0.934	-1.204	-0.669
	Size D	0.414	0.059	0.739
	Size E	-0.218	-0.660	0.191
	Price	-0.904	-1.472	-0.309
	Promotion	-2.714	-4.229	-0.894
Market 3	Size A	0.749	0.397	1.093
	Size B	0.237	-0.087	0.430
	Size C	-0.594	-1.015	-0.224
	Size D	-0.641	-0.877	-0.367
	Size E	-0.545	-0.821	-0.271
	Price	-0.593	-0.924	-0.211
	Promotion	-3.388	-4.764	-2.079

Note: HPDR stands for Highest Posterior Density Region.

Table 6: Application: Posterior Mean and HPDR of the Fixed Elements of f_m (size and price and promotion coefficients) for 3 out 18 markets

	Posterior	HPDR	
	Mean	1%	99%
D_1^2	1.436	0.801	2.493
D_2^2	0.798	0.653	0.896
D_3^2	1.341	1.184	1.619
D_4^2	0.831	0.628	1.461
D_5^2	0.293	0.249	0.335
D_6^2	0.874	0.739	1.022
D_7^2	1.853	0.970	4.412
D_8^2	0.500	0.443	0.575
D_9^2	4.648	1.509	9.801
D_{10}^2	0.395	0.327	0.457
D_{11}^2	1.197	0.787	1.718
D_{12}^2	0.569	0.466	0.727
D_{13}^2	0.556	0.478	0.603
D_{14}^2	0.628	0.522	0.700
D_{15}^2	10.625	3.911	21.916
D_{16}^2	0.432	0.329	0.502
D_{17}^2	0.195	0.160	0.238
D_{18}^2	0.990	0.750	1.257
D_{19}^2	3.249	1.824	5.073
D_{20}^2	0.221	0.182	0.276
D_{21}^2	0.337	0.248	0.378
D_{22}^2	0.602	0.540	0.698
D_{23}^2	7.294	5.028	12.136
D_{24}^2	0.700	0.623	0.794
D_{25}^2	2.361	1.697	2.710
D_{26}^2	0.754	0.654	0.955
D_{27}^2	0.643	0.522	0.732
D_{28}^2	0.558	0.472	0.667
D_{29}^2	0.579	0.468	0.672
D_{30}^2	0.587	0.474	0.797
D_{31}^2	0.590	0.443	0.777
D_{32}^2	0.624	0.490	0.811

Note: HPDR stands for Highest Posterior Density Region.

Table 7: Application: Posterior Distribution of the Elements of the D^2 matrix, where $\Psi = DSD$.

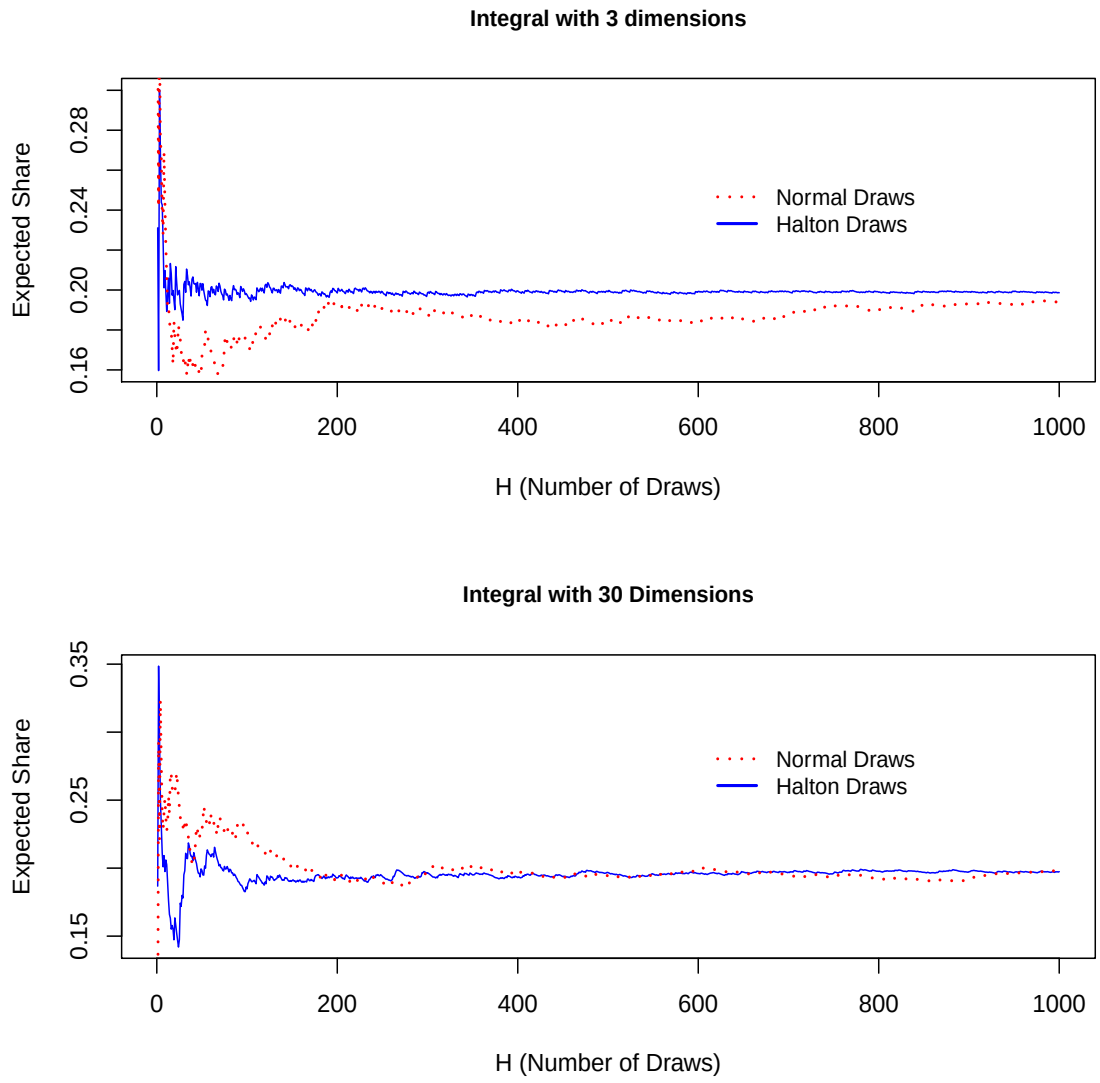

Figure 1: Performance of Halton Based Normal Draws versus Normal Draws

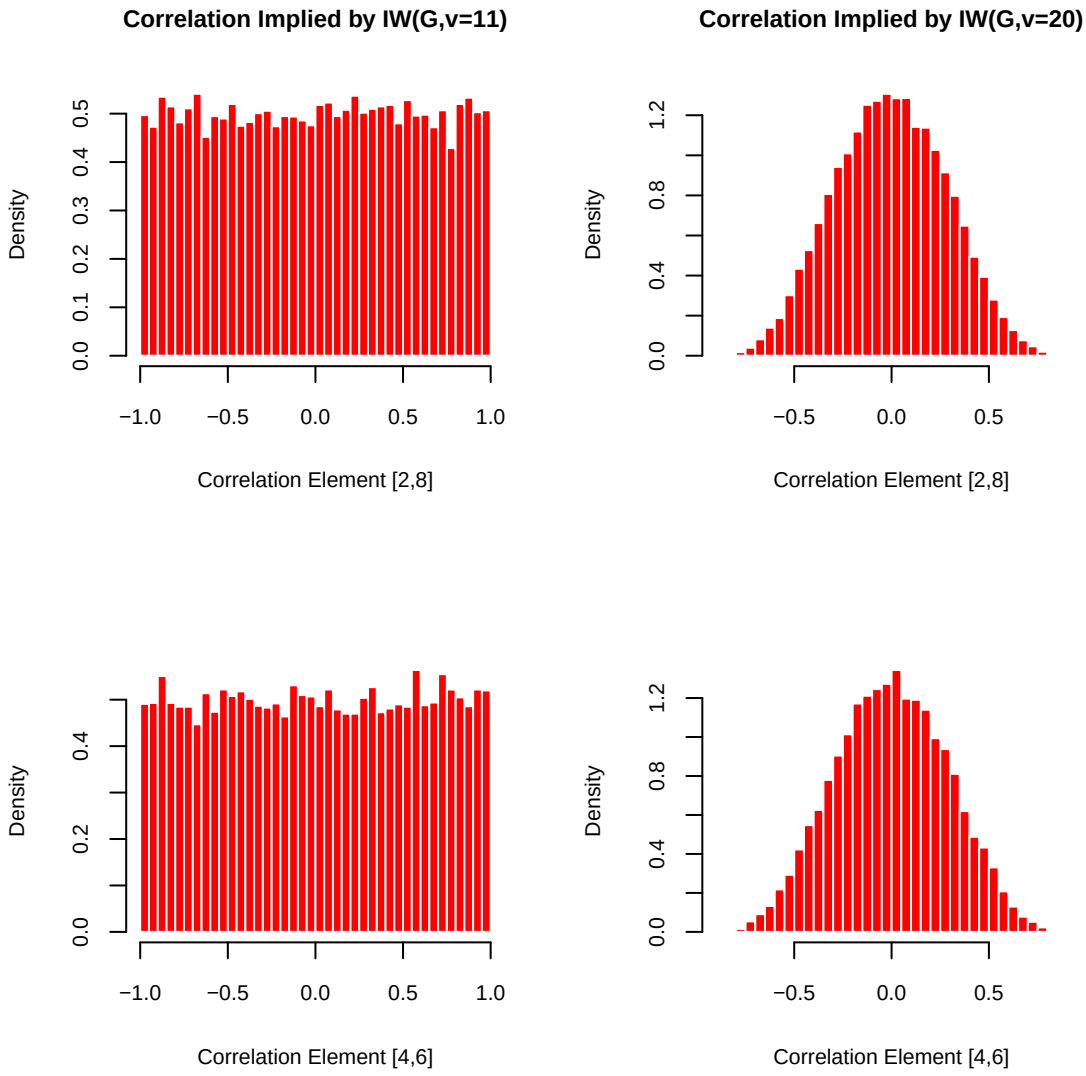

Figure 2: Prior Correlations for Different Elements of Ψ. The degrees of freedom for the Wishart Distribution v are set to 11 for the left panel and 20 for the right panel.

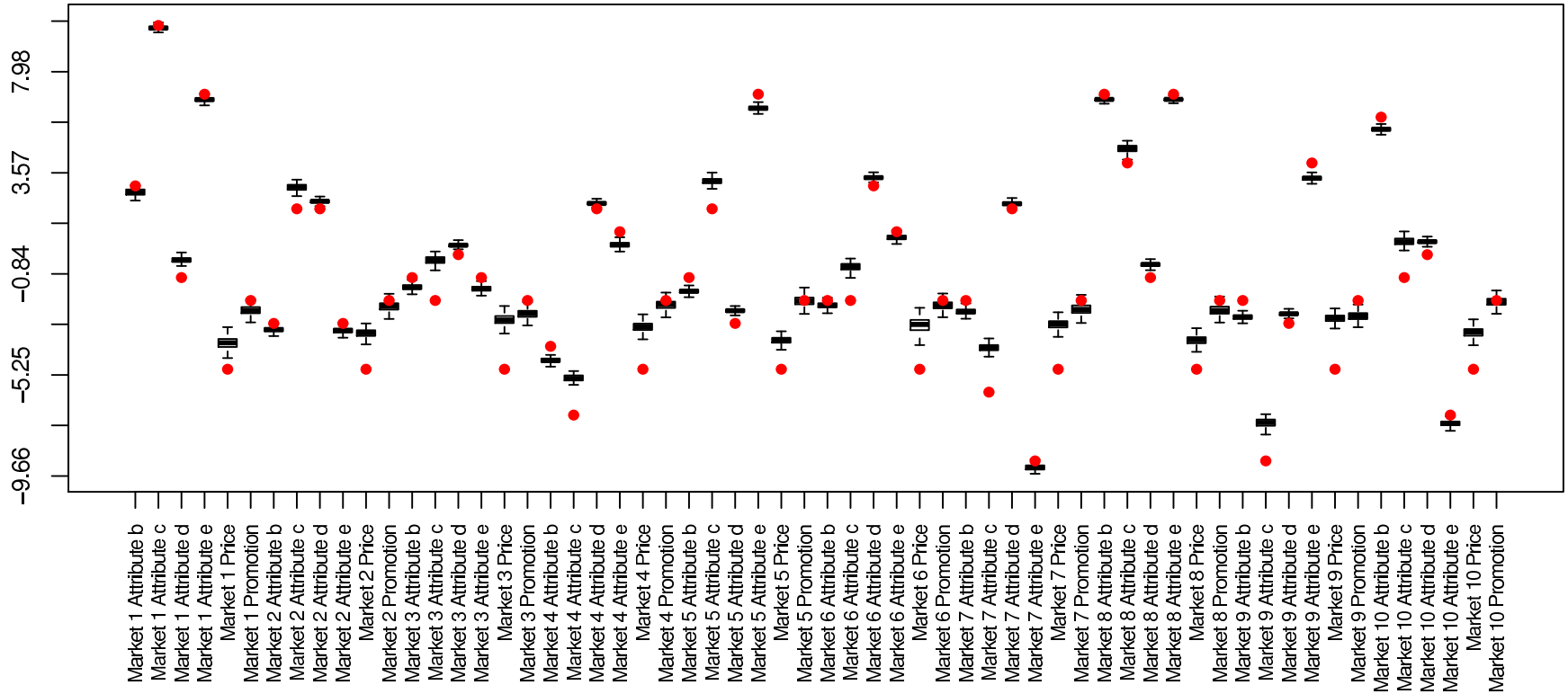

Figure 3: Simulation Experiment: Real (Circles) versus the Posterior Distribution (Box-plots) of the Fixed Coefficients

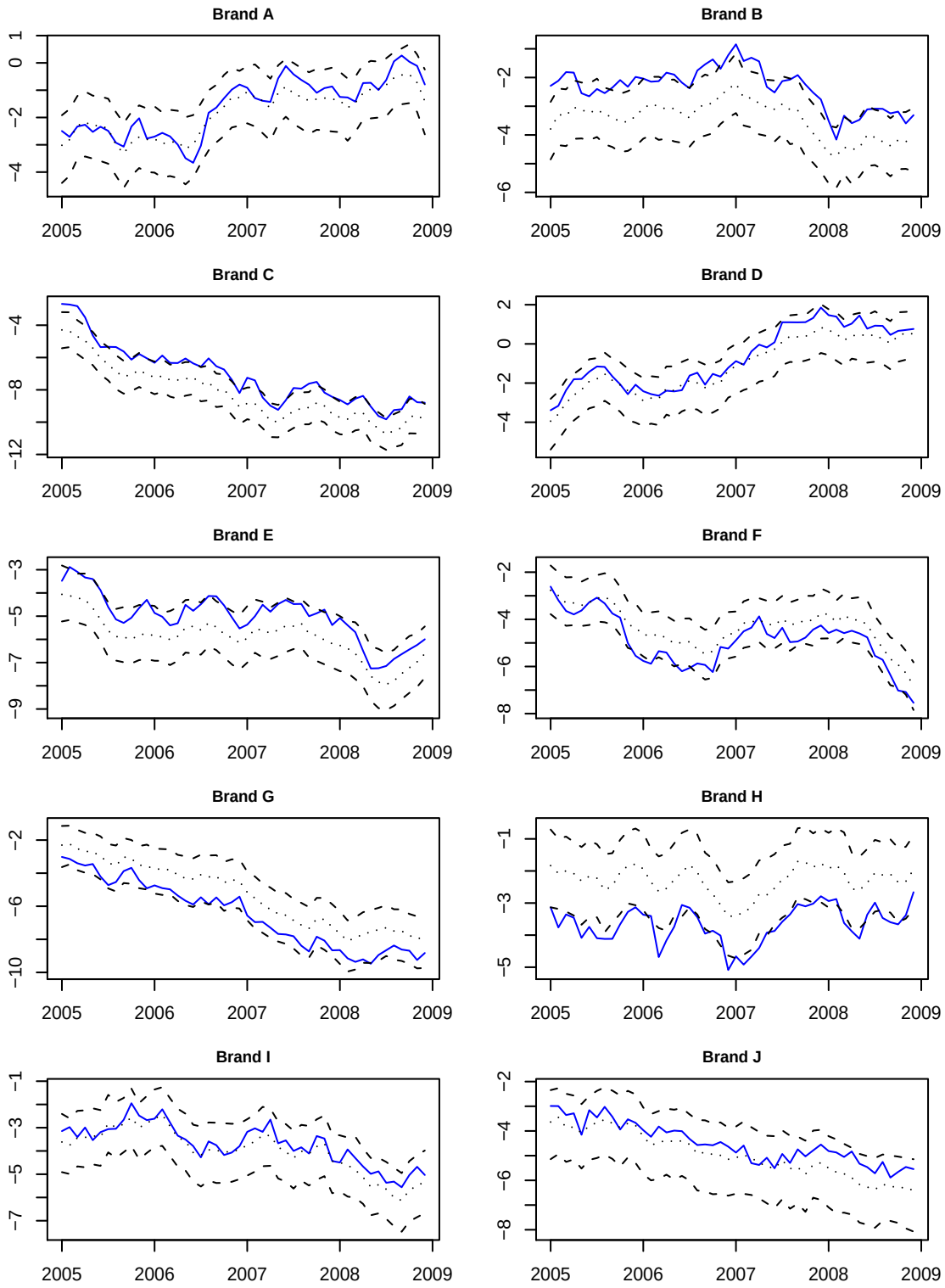

Figure 4: Simulation Experiment: Real (Solid Line) versus Posterior Mean (Dots) and the 99% HPDR (Dashed Lines) of the Time-Varying Brand Coefficients at Market 5

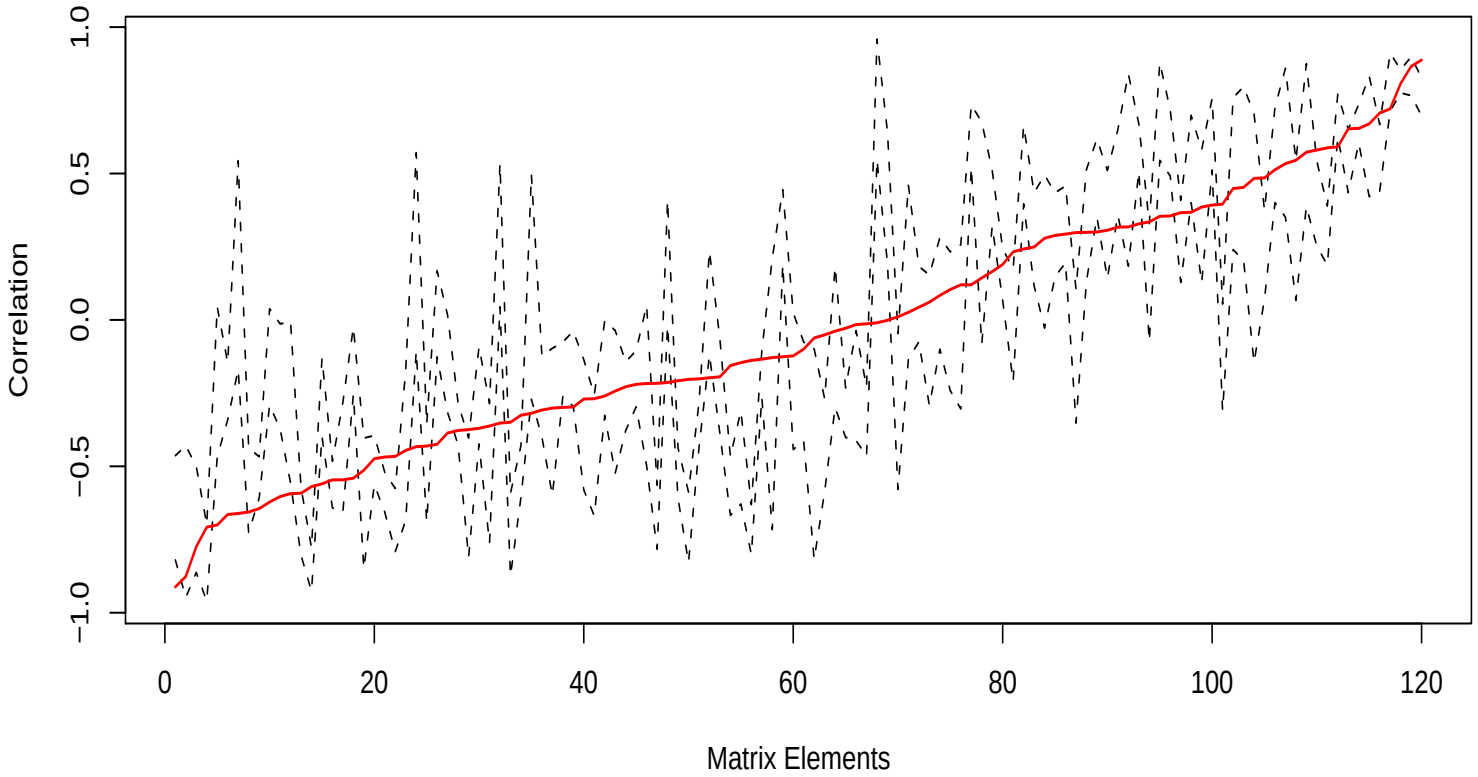

Figure 5: Simulation Experiment: Real (Solid Line) versus Posterior 99% HPDR (Dashed Lines) of All Elements in the Correlation Matrix R.

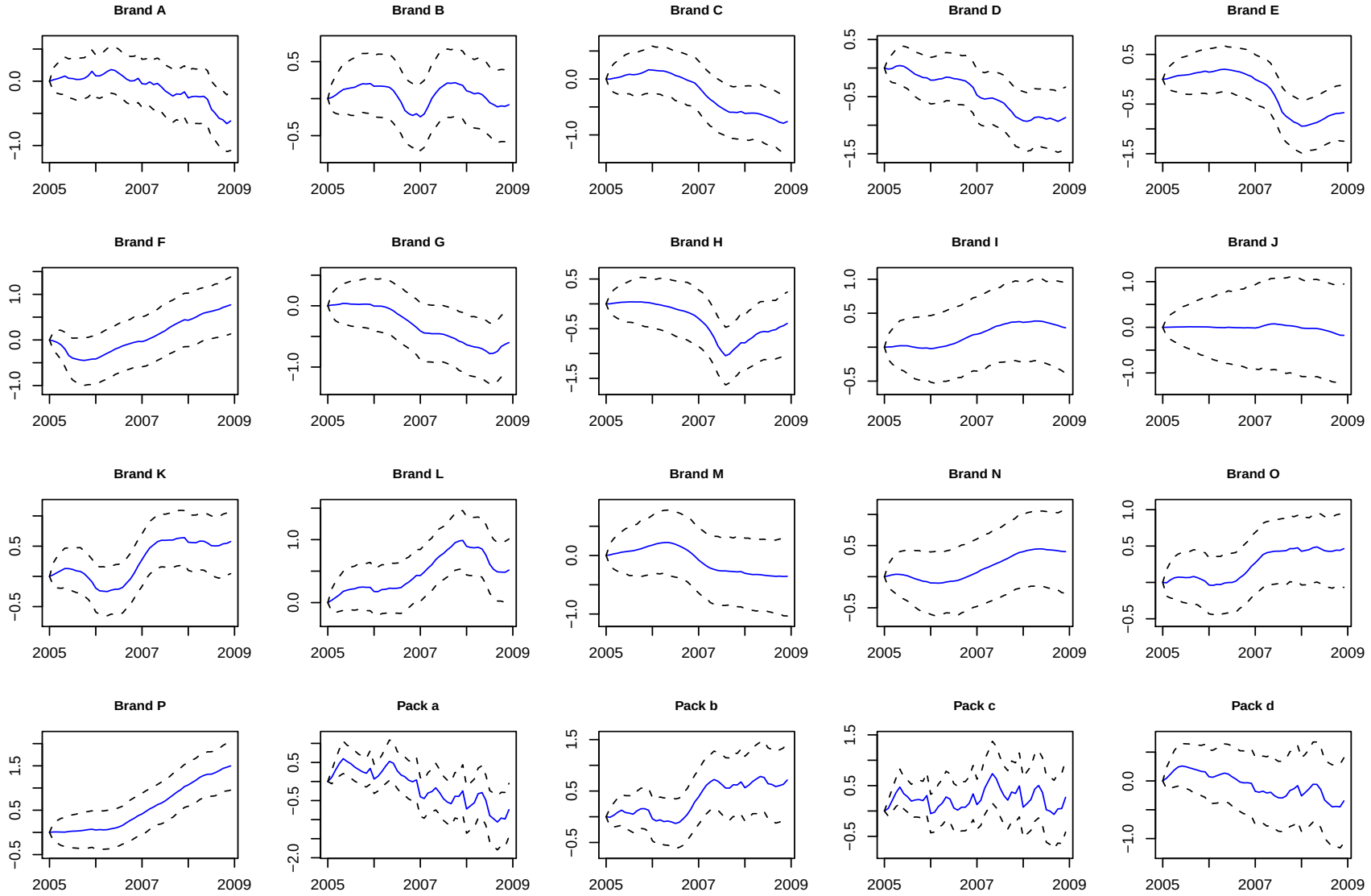

Figure 6: Application: Time-Profile Relative to First Period of 8th Market Time-Varying Factors f^m (Solid Lines) and their 99% HPDR (Dashed Lines).

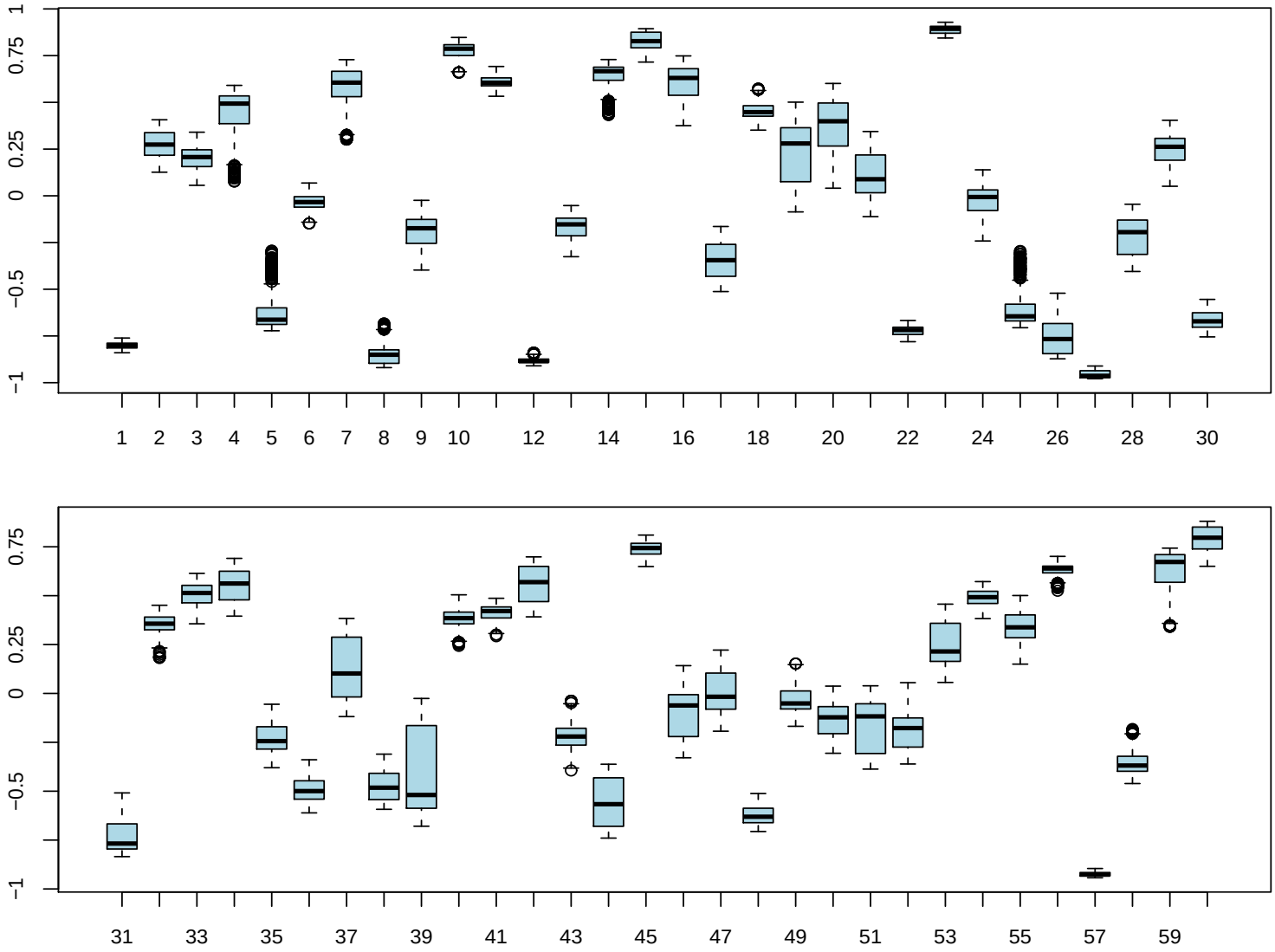

Figure 7: Application: Distribution of 60 Correlation Elements of the Ψ matrix size 32×32

Figure 8: Own Price Elasticity for Products at market 2

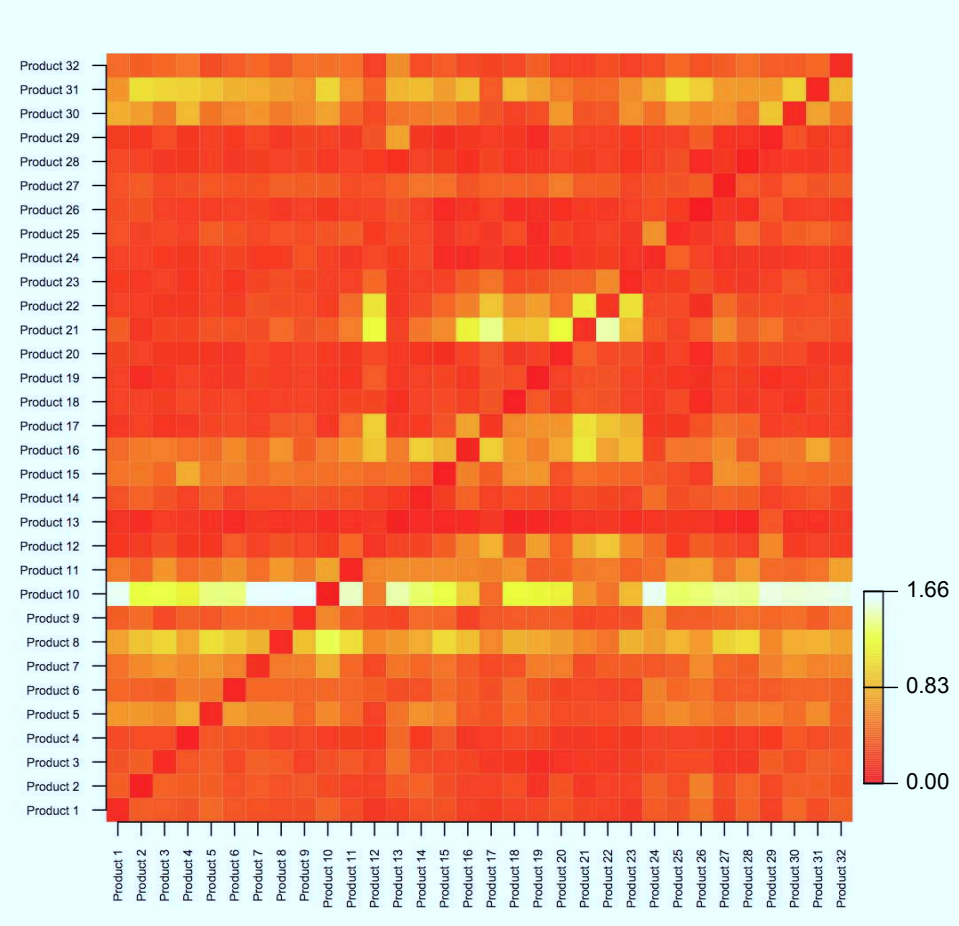

Figure 9: Cross-Price Elasticities at Market 2

A Appendix

Iterative BLP Procedure

We use the iterative procedure proposed in BLP to obtain the $\mu_t = (\mu_{1t}, \ldots, \mu_{Jt})$. Note that, for convenience, we omit the market indicators m here. The procedure consists of the following steps. First we obtain H draws of υ_i. To this end we write the d-th draw as $\upsilon_i^d = \Psi^{1/2}\zeta^d$ where ζ^d is draw from a joint normal distribution which we obtain using scrambled Halton draws. Given these draws and some initial value for μ_t we can compute the implied market shares $\hat{s}_t$. Given the shares, both real s_t and implied $\hat{s}_t$, we can use the contraction mapping

$$\mu_t^{new} = \mu_t^{old} + \log(s_t) - \log(\hat{s}_t) \tag{26}$$

to obtain a new value for μ_t. We repeat the contraction mapping computing the implied shares $\hat{s}_{jt}$ as

$$\left(\sum_{i=1}^{H} \frac{\exp(\mu_{jt}^{old} + \lambda_{jt}' \upsilon_i)}{\exp(\mu_{0t}) + \sum_k \exp(\mu_{kt}^{old} + \lambda_{kt}' \upsilon_i)} \right) / H \tag{27}$$

for $j = 1, \ldots, J$ and $t = 1, \ldots, T$ and we stop the contraction mapping when the values of μ_t^{new} and μ_t^{old} converge.

Note that we include and solve for the outside good utility μ_{0t} in the contraction mapping iterations. We discovered that the precision of the contraction mapping is higher when we iterate over the utilities of all products together with the utility of the outside good.

Computing Elasticities

We use the following definition to compute the price elasticities φ_{jl}^m between product j and l in market m conditional on all model parameters:

$$\varphi_{jl}^m = \frac{p_{lt}^m}{\mathrm{E}[s_{jt}^m]} \frac{\partial \mathrm{E}[s_{jt}^m]}{\partial p_{lt}^m} = \frac{p_{lt}^m}{\int s_{jt}^m \pi(\eta_t; 0, \tau_m^2 \mathbf{I}) d\eta_t^m} \times \int \frac{\partial}{\partial p_{lt}^m} s_{jt}^m \pi(\eta_t | \tau_m^2) d\eta_t^m \tag{28}$$

where s_{jt}^m is defined in (7) and

$$\frac{\partial}{\partial p_{lt}^m} s_{jt}^m = \begin{cases} -\beta^m \int s_{ijt}^m s_{ikt}^m \phi(\upsilon_i^m; 0, A^m \Psi A_m') d\upsilon_i^m & \text{if } k \neq j \\ \beta^m \int s_{ijt}^m (1 - s_{ikt}^m) \phi(\upsilon_i^m; 0, A^m \Psi A_m') d\upsilon_i^m & \text{if } k = j, \end{cases} \tag{29}$$

and β^m is the price coefficient in market m. Finally, to obtain the posterior distribution of the price elasticities we average (28) over the posterior draws for all parameters.

References

Barnard, J., McCulloch, R., Meng, X.-L., 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica 10 (4), 1281–1312.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Berry, S., Pakes, A., 2007. The pure characteristics demand model. International Economic Review 48 (4), 1193–1225.

Bodapati, A. V., Gupta, S., 2004. The recoverability of segmentation structure from store-level aggregate data. Journal of Marketing Research 41 (3), 351–364.

Chintagunta, P., Dube, J.-P., Goh, K. Y., 2005. Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. Management Science 51 (5), 832–849.

Durbin, J., Koopman, S., 2002. A simple and efficient simulation smoother for state space time series analysis. Biometrika 89 (3), 603–615.

Fruhwirth-Schnatter, S., 2004. Efficient bayesian parameter estimation. In: Harvey, A., Koopman, S., Shephard, N. (Eds.), State Space and Unobserved Component Models: Theory and Applications. Press Syndicate and the University of Cambridge, Ch. 7, pp. 123–151.

Goeree, M. S., 2008. Limited information and advertising in the us personal computer industry. Econometrica 76 (5), 1017–1074.

Haran, M., Hodges, J. S., Carlin, B. P., 2003. Accelerating computation in markov random field models for spatial data via structured mcmc. Journal of Computational and Graphical Statistics 12 (2), 249–264.

Jiang, R., Manchanda, P., Rossi, P. E., 2009. Bayesian analysis of random coefficient logit models using aggregate data. Journal of Econometrics 149 (2), 136–148.

Musalem, A., Bradlow, E. T., Raju, J. S., 2006. Bayesian estimation of random-coefficients choice models using aggregate data. Journal of Applied Econometrics 24 (3), 490–516.

Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica 69 (2), 307–342.

Park, S., Gupta, S., 2009. Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. Journal of Marketing Research 46 (4), 531–542.

Train, K., 2003. Discrete choice models with simulation. Cambridge University Press.

Yang, S., Chen, Y., Allenby, G. M., 2003. Bayesian analysis of simultaneous demand and supply. Quantitative Marketing and Economics 1 (3), 251–275.

Publications in the Report Series Research* in Management

ERIM Research Program: "Marketing"

2010

Deviation Among Technology Reviews: An Informative Enrichment of Technology Evolution Theory for Marketing
Ashish Sood and Stefan Stremersch
ERS-2010-005-MKT
http://hdl.handle.net/1765/17766

Perceptual maps: the good, the bad and the ugly
John Gower, Patrick Groenen, Michel Van de Velden and Karen Vines
ERS-2010-011-MKT
http://hdl.handle.net/1765/18462

Do Charities Get More when They Ask More Often? Evidence from a Unique Field Experiment
Bas Donkers, Merel van Diepen, and Philip Hans Franses
ERS-2010-015-MKT
http://hdl.handle.net/1765/19423

Random Coefficient Logit Model for Large Datasets
Carlos Hernández Mireles, and Dennis Fok
ERS-2010-021-MKT
http://hdl.handle.net/1765/19669

The Launch Timing of New and Dominant Multigeneration Technologies
Carlos Hernández Mireles, and Philip Hans Franses
ERS-2010-022-MKT
http://hdl.handle.net/1765/19670

Finding the Influentials that Drive the Diffusion of New Technologies
Carlos Hernández Mireles
ERS-2010-023-MKT
http://hdl.handle.net/1765/19671

---

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship