

# Bayesian Estimation of Random-Coefficients Choice Models

## Using Aggregate Data

Andrés Musalem

Eric T. Bradlow

Jagmohan S. Raju\*

July 2007

---

\*Andrés Musalem is Assistant Professor of Marketing at The Fuqua School of Business of Duke University. Eric T. Bradlow is The K.P. Chao Professor, Professor of Marketing, Statistics and Education and Academic Director of the Wharton Small Business Development Center, The Wharton School of the University of Pennsylvania. Jagmohan S. Raju is The Joseph J. Aresty Professor, Professor of Marketing, The Wharton School of the University of Pennsylvania. The authors thank Yuxin Chen, Sha Yang, Charles J. Romeo, Pradeep Chintagunta, Peter Rossi, Lawrence D. Brown, Edward I. George, David C. Schmittlein and David R. Bell for their valuable comments. In addition, we thank the editor and two anonymous reviewers for their constructive suggestions that have greatly improved this manuscript. Please address all correspondence to Andrés Musalem, 1 Towerview Drive, Durham, NC 27708, Phone: (919) 660-7827, Fax: (919) 660-7990, amusalem@duke.edu.

# Bayesian Estimation of Random-Coefficients Choice Models

## Using Aggregate Data

### Abstract

This article discusses the use of Bayesian methods for estimating logit demand models using *aggregate* data. We analyze two different demand systems: independent samples and consumer panel. Under the first system, there is a different and independent random sample of  $N$  consumers in each period and each consumer makes only a single purchase decision. Under the second system, the same  $N$  consumers make a purchase decision in each of  $T$  periods. Interestingly, there exists an asymptotic link between these two systems, which has important implications for the estimation of these demand models. The proposed methods are illustrated using simulated and real data.

*Key Words:* *Discrete Choice Models, Aggregate Data, Bayesian Methods, Markov Chain Monte Carlo Simulation, Data Augmentation, Random Coefficients.*

# 1 Introduction

During the last two decades, many researchers in Marketing and Economics have relied on aggregate data to make inferences about consumer preferences and the strategic behavior of firms. These aggregate data typically contain information solely on how many consumers chose each brand in each period without the knowledge of the choices of each individual consumer. Such analyses are commonplace both for practical reasons and for academic reasons as a result of data ubiquity in this form - aggregate information is typically cheaper and easier to collect or acquire. Some examples of applications using this kind of information include empirical studies based on supermarket scanner data (e.g., Christen *et al.*, 1997; Besanko *et al.*, 2003; Berto-Villas Boas, 2007), airline data (e.g., Berry *et al.*, 2006) and automobile sales data (e.g., Berry *et al.*, 1995; Sudhir, 2001), to name just a few.

From a methodological point of view, since consumers are commonly assumed to be heterogeneous in terms of their preferences, this heterogeneity must be taken into account in order to obtain unbiased estimates of the demand function (Kim, 1995; Chintagunta, 2001) and, consequently, to draw better inferences about interfirm competition (e.g., Berry *et al.*, 1995; Nevo, 2001; Berto-Villas Boas, 2007). When disaggregate data are available (e.g., Rossi *et al.*, 1996), it is straightforward to account for this heterogeneity by using Bayesian methods or Maximum Simulated Likelihood estimation (Huber and Train, 2001). However, when dealing with aggregate information, yet still wanting to incorporate heterogeneity, some difficulties arise that have led researchers to simplify the formulation of the likelihood function (e.g., Kim, 1995) or to use non likelihood-based methods (e.g., Berry, 1994).

In this article, we discuss the use of Bayesian methods normally “reserved” for data that arrive in the form of individual-level choices, for estimating demand models from aggregate market share data. An important property of the demand models under study is that they are formulated as the aggregation of the individual choices of the consumers that generated the aggregate data. From a theoretical point of view, this implies that the demand model can be constructed by specifying individual-level assumptions about the behavior of consumers in the market (e.g., utility

maximization), which provides a theoretical justification for the estimation of the demand models under study.

Moreover, from a practical point of view, the use of individual-level assumptions about consumer behavior enables us to estimate the distribution of consumer preferences, which in turn provides valuable information to the researcher. For example, using this information it is possible to estimate the extent to which consumers switch brands, the proportion of heavy users of a given brand and the penetration of each product in the market; and *yes* all from aggregate data. Similarly, the consequences of policy experiments can also be simulated, such as determining how many consumers would switch from a given brand to another if prices were temporarily reduced.

In addition, the use of Bayesian methods as done here, offers several other potential benefits. First, finite-sample inferences about the parameters of the model can be directly obtained without resorting to asymptotics unnecessarily or when they are unlikely to apply. Second, it is straightforward to derive finite-sample inferences for any quantity of interest that can be computed as a function of the model parameters and the data (e.g., penetration, brand switching, cross-price elasticities, market shares, profits, consumer welfare). Third, Bayesian methods provide a conceptually simple approach for dealing with complex models containing multiple parameters (Gelman *et al.*, 1995). Fourth, the specification of a prior distribution for the parameters of the model allows the researcher to incorporate auxiliary information (e.g., expert knowledge) into the estimation in a coherent fashion (e.g., Bajari and Ye, 2003).

In the context of the statistical problem analyzed here, the estimation of the distribution of consumer preferences depends crucially on the assumptions about the demand system, more so than when individual-level data are available and the adequacy of the assumptions can be more easily checked. In particular, one could consider the following two alternative systems of demand which are based on two different assumptions:

1. Independent Samples: Assume that for each of  $T$  periods there is a different random sample of  $N$  consumers that make purchase decisions in the market. These consumers are randomly drawn from an *infinite* population. For example, assuming  $N = T = 2$ , in the first period we could observe Consumers 1 and 7 choosing alternatives A and B, respectively, while in

period 2, we might observe Consumers 3 and 8 choosing alternatives A and C, respectively (where the set of alternatives may include a no-purchase option). Therefore, there are  $T$  different (independent) random samples of consumers and for each of them we only have one (aggregate) observation.

2. Consumer Panel: Assume there are  $N$  consumers in the market and each of these consumers makes a purchase decision in each of  $T$  periods. For example, Consumers 1 and 2 might choose products A and B in the first period and then, in the second period, the same Consumers 1 and 2 might choose products A and C, respectively. Therefore, this system corresponds to a panel of consumers for which  $T$  (aggregate) observations are available.

In real applications, however, a typical market may not strictly correspond to either one of these two extreme cases but instead to a combination of them, where each consumer makes purchase decisions in some of the  $T$  periods. Nevertheless, as we discuss later in §5, the methods developed for the estimation of these two extreme systems of demand can also be applied for the estimation of this hybrid scenario. For expositional purposes, we focus only on these two extreme cases.

In the case of the first system (independent samples), Bayesian estimation can be implemented using Markov Chain Monte Carlo (MCMC) methods that are very similar in terms of complexity and computational requirements to those used when disaggregate data are available (e.g., Yang *et al.*, 2003). Moreover, for this system of demand, it is possible to carry out the estimation by modeling the latent preferences of only a subsample of  $R < N$  consumers in each period, leading to great computational savings. In the case of the second system (panel of consumers), we demonstrate that Bayesian estimation of the distribution of consumer preferences can be implemented by augmenting the aggregate data (Tanner and Wong, 1987) with an unobserved sequence of individual choices for each of the  $N$  consumers in the panel. It is shown here that these sequences of choices can be sampled from their posterior distribution directly using a Gibbs step: a complementary yet different approach to recent work (Chen and Yang, 2007) that utilizes a Metropolis-Hastings (MH) step.

Interestingly, it can be shown by applying the Weak Law of Large Numbers (WLLN) that, as

the number of consumers making purchase decisions in each period increases, the observed demand for each of these systems (independent samples and consumer panel) converges in probability to the same values. Moreover, the information obtained about the demand parameters is asymptotically equivalent under both demand systems when  $N$  approaches infinity. This implies that the methods used for the estimation of the system of independent consumers can be used to estimate the demand of consumers from the second system (panel of consumers). This, in turn, provides substantial computational benefits that are particularly relevant for practical applications where  $N$ , the number of consumers in a market, is likely to be very large, say  $N=100,000$  or more.

The rest of the paper is organized as follows. In Section 2 we introduce the demand model for the first system (independent samples) and discuss the estimation of this system via MCMC simulation. In Section 3 we describe the model and the estimation for the second system of demand (panel of consumers). In Section 4, we discuss the asymptotic link between these two systems and the computational implications for the estimation of the second system. In Section 5 we discuss model selection issues. In Section 6 we apply the proposed methods to estimate the distribution of preferences of consumers using a data set containing information on purchases and marketing activity in the facial tissue product category. Finally, Section 7 concludes this article with general recommendations and conclusions for conducting Bayesian inference in similar contexts.

## 2 Independent Samples Case

In this section, we present the model of demand and the estimation procedure for the case where the aggregate data for each period are generated by a different sample of consumers which are randomly drawn from an infinite population. We note that this assumption is similar to the one used by Bodapati and Gupta (2004) for the derivation of their likelihood function.

### 2.1 Demand Model

In the case of the first system of demand, we assume that for each period  $t = 1, \dots, T$  there is a different random sample of  $i = 1, \dots, N$  consumers that make purchase decisions. Denoting by  $y_{it}$

the choice of the  $i^{\text{th}}$  consumer in the sample from period  $t$ ,  $y_{it}$  satisfies:

$$\begin{aligned}
y_{it} &= \operatorname{argmax}_j U_{ijt} \\
&= \operatorname{argmax}_j V_{ijt} + \epsilon_{ijt} \\
&= \operatorname{argmax}_j \theta'_{it}x_{jt} + \xi_{jt} + \epsilon_{ijt},
\end{aligned} \tag{1}$$

where  $U_{ijt}$  is the utility of alternative  $j$  for the  $i^{\text{th}}$  consumer in the  $t^{\text{th}}$  period random sample;  $V_{ijt}$  is defined as the sum of  $\theta'_{it}x_{jt}$  and  $\xi_{jt}$ , the deterministic component of utility;  $\theta_{it}$  is a vector of coefficients for the  $i^{\text{th}}$  consumer in period  $t$ ;  $x_{jt}$  is a vector of attributes for brand  $j$  in period  $t$ , e.g., including price, brand dummies and other product characteristics;  $\xi_{jt}$  is a common demand shock that affects the utility of alternative  $j$  in period  $t$  for all consumers; and,  $\epsilon_{ijt}$  is an individual-specific demand shock for the utility of alternative  $j$  for the  $i^{\text{th}}$  consumer in period  $t$ .

Assuming  $\epsilon_{ijt}$  is distributed according to an Extreme Value (0,1) distribution, the probability,  $p_{ijt}$ , that the  $i^{\text{th}}$  consumer in period  $t$  chooses brand  $j$  is given by (Ben Akiva and Lerman, 1985):

$$p_{ijt} = \frac{e^{V_{ijt}}}{\sum_{k=1}^J e^{V_{ikt}}}, \tag{2}$$

where  $J$  is the number of alternatives that the consumer is considering. We note that one of these alternatives may correspond to a no-purchase option, as in the models estimated in §4, §5 and §6.

From a statistical point of view, the main goal is to estimate the distribution of preferences of the consumers that generated the aggregate data. These preferences are governed by the distribution of the coefficients  $\theta_{it}$ . As in many applications (e.g., Berry *et al.*, 1995), we add a parametric assumption by specifying that the coefficients  $\theta_{it}$  are i.i.d. Multivariate Normal with mean  $\bar{\theta}$  and variance-covariance matrix  $D$ . Therefore, our goal is to estimate  $\bar{\theta}$  and  $D$ , or in the Bayesian sense, provide valid inferences for their posterior distributions based solely on aggregate data, i.e. the observed market share of each alternative  $j$  in each period  $t$ , denoted  $S_{jt}$ .

In addition, it is necessary to note that the inclusion of the common demand shocks ( $\xi_{jt}$ ) enables us to capture shifts in demand due to factors that are unobserved to the researcher. This

is particularly important when the number of consumers in the market is large (say,  $N=100,000$ ). In those cases, if the common demand shocks are not included, then by the Weak Law of Large Numbers (see §4) the model becomes nearly deterministic in the sense that each of the observed market shares in a given period is almost equal to a deterministic function of  $\{x_{jt}\}_{j=1}^J$ ,  $\bar{\theta}$  and  $D$ . This implies, for example, that if all marketing variables (e.g., prices, feature and display) take the same values in two different time periods, we should approximately observe the same market shares in both periods. Consequently, if there are reasons to believe that market shares may change across time periods by a larger magnitude, then this can be handled by including and estimating these common demand shocks<sup>1</sup>. The estimation of these common demand shocks requires some parametric assumptions, therefore, we will assume that each  $J$ -dimensional vector of brand demand shocks ( $\xi_t$ ) is i.i.d. Multivariate Normal with a zero mean and variance-covariance matrix  $\Sigma$ . Finally, we note that these demand shocks may be correlated with some of the covariates in the utility function.

## 2.2 Parameter Interpretation and Identification

We note that the each of the elements of  $\bar{\theta}$ ,  $D$  and  $\Sigma$ , the parameters that govern our model, has an intuitive interpretation and, in terms of identification, the values of each of them induce different patterns on the aggregate data, hence the model can be identified. In what follows, we focus our discussion on the elements of  $D$  and  $\Sigma$ , the variance-covariance matrices of the demand coefficients and common demand shocks, and their relationship to  $\bar{\theta}$  in some instances<sup>2</sup>.

The diagonal elements of  $D$  measure the extent to which there is heterogeneity among consumers in terms of their preferences. For example, a large variance of the intercepts of a given brand implies that after controlling for all other covariates (e.g., prices, promotional activity, demand shocks), there are important differences in terms of the baseline preferences of consumers for that brand. This is the case, for instance, when there is a substantial number of consumers that are very loyal to the brand, while there is a large number of consumers who never buy that brand.

---

<sup>1</sup>We thank an anonymous reviewer for highlighting the importance of including common demand shocks in the utility function to solve this problem.

<sup>2</sup>Note that similar conclusions hold for the case of a panel of consumers and, hence, are not described in §3.



Consequently, the market shares of this brand will exhibit a small number of fluctuations when some of the other explanatory variables (prices, promotional activity, demand shocks) change across time periods because these changes are less effective at shifting consumer preferences for a given brand.

The variances of the preference coefficients associated with prices and other covariates (the non-brand elements of  $D$ ) also have an intuitive interpretation. A large variance in price coefficients will be associated with the existence of a group of customers that strongly react to price changes and another group of customers whose choices are almost unaffected by price movements. The effect of this parameter on the observed market share data also depends on the mean of the price coefficients. For example, if the mean of the price coefficient across consumers equals -1 ( $\bar{\theta}_{\text{price}} = -1$ ), then as the variance of these coefficients ( $D_{\text{price,price}}$ ) increases (e.g., from 0.1 to 3.0), prices start exhibiting a positive effect on the utilities of some consumers, compensating the negative effect of prices on the utility of the remaining consumers. Therefore, fluctuations in prices have a more limited impact on the shares of each of the (purchase) alternatives and, consequently, the variance of the market shares across periods decreases.

In terms of the off-diagonal elements of  $D$ , the covariance between the brand intercepts of two different brands measures the extent to which preferences for the two brands are linked. For example, a positive covariance between the brand intercepts for Brand 1 and Brand 2 implies that if a consumer exhibits a strong intrinsic preference for Brand 1, then it is likely that the consumer also has a strong preference for Brand 2. As this covariance increases, the utilities of Brand 1 and Brand 2 become more similar in all periods and it is, therefore, “easier” for the utility of the other alternatives to be larger than the maximum between the utilities of Brand 1 and Brand 2. As a consequence, the average share of all other brands increases.

In the case of the covariance between price coefficients and the intercepts for a given brand, say Brand 1, a negative value of this parameter indicates that those consumers that exhibit higher intrinsic preferences for that brand are also likely to be more price elastic. As this negative covariance increases, consumers with a stronger intrinsic preference for Brand 1 are on average less price sensitive, therefore, prices are less likely to drive customers to switch from Brand 1

to other brands, and, thus the market shares of the remaining brands become more interrelated (e.g., the changes in the share of Brand 2 have more to do with gaining or losing customers from/to Brand 3), consequently, the magnitude of the correlation between the market shares of other brands increases.

In terms of the covariance matrix of the common demand shocks ( $\Sigma$ ), the diagonal elements of this matrix measure the extent to which the utilities of *all* consumers fluctuate from period to period for reasons other than changes in the value of the explanatory variables ( $x_t$ ). In particular, a higher value of  $\Sigma_{jj}$  would be associated with higher fluctuations of the market share of Brand  $j$  across time periods (higher  $\text{var}(S_{jt})$ ). In the case of the off-diagonal elements of this matrix, when, for example,  $\Sigma_{12}$  increases, the utilities of Brand 1 and Brand 2 become more similar, therefore the market shares of the remaining alternatives increase (given utility-maximizing consumers). In addition, as this covariance increases, when the utility for Brand 1 increases (or decreases), there is more likely to be a similar change for the utility of Brand 2, therefore demand shocks are more likely to affect the share of the remaining alternatives (i.e., if the demand shocks for both brands are simultaneously positive or negative, the remaining alternatives may lose or gain share to/from both Brand 1 or Brand 2). Consequently, the share of these other alternatives depends more on the share of Brand 1 and Brand 2 (i.e., stronger correlation between the market shares of the remaining brands and the market shares of Brand 1 and Brand 2).

Finally, extensive simulations (available upon request) using a full-factorial design for  $\bar{\theta}$ ,  $D$  and  $\Sigma$  demonstrate these effects and show that variations in each of the demand parameters generate different patterns of market shares, even after controlling for variations of other demand parameters. Furthermore, we believe that future research should be aimed at complementing this numerical analysis and the simulation experiments presented in this paper with a theoretical derivation of the conditions that ensure identification of the parameters of the model (see Berry *et al.*, 2004, and Bodapati and Gupta, 2004, for related theoretical results).

### 2.3 Estimation for the Independent Samples Case

In terms of the estimation procedure, which is implemented via Gibbs sampling, we augment the data (observed market shares) by treating the unobserved individual choices as parameters (Chen and Yang, 2007). For ease of exposition, we will assume in the remainder of this section and in the next one that each of the demand shocks is equal to zero (i.e.,  $\xi_{jt} = 0$ ), and then, in §4, where we analyze the case in which  $N$  is large, we will explicitly address the computational issues associated with the estimation of these common demand shocks.

For the model analyzed in this section (independent samples), every consumer makes a purchase decision in no more than one period. Thus, the indices assigned to consumers choosing each brand in each period are completely arbitrary, since we can reassign these indices without affecting the posterior distribution of  $\bar{\theta}$  and  $D$  as there is no linkage of choices across time. For example, suppose that consumers 3 and 7 appear in the random sample corresponding to period  $t$  and that they choose brand  $A$  and  $B$ , respectively, in that period. Then, our inferences about  $\bar{\theta}$  and  $D$  would be exactly the same if these choices were interchanged, i.e., if consumer 3 chose  $B$  and consumer 7 chose  $A$ . This corresponds to a label switching problem that is well-studied (e.g., Stephens, 2000) and it can be alleviated as follows. For each period  $t$ , we can arbitrarily and without any loss of generality assign the first  $NS_{1t}$  indices to those consumers who chose brand 1; the next  $NS_{2t}$  indices to those who chose brand 2, and so on. These indices remain fixed at all iterations of the Gibbs sampler, which makes the estimation problem similar in terms of computational requirements to the corresponding one when disaggregate data are available and improves the stability of the Markov Chain by reducing the label switching problem. In contrast, we note that in the consumer panel case, these indices play a critical role and it is necessary to simulate the sequence of choices of each consumer from its posterior distribution. The simulation of these sequences will be described in detail in §3.

In what follows, we introduce additional notation and we formulate the augmented likelihood and posterior density of the parameters. Let  $z_{ijt}$  be equal to 1 if the  $i^{\text{th}}$  consumer in period  $t$  chooses alternative  $j$ , otherwise  $z_{ijt}$  is equal to zero. For notational convenience, we denote by  $Z$  the matrix containing each of the elements  $z_{ijt}$ . Accordingly, the likelihood of the augmented

data ( $Z$ ) can be computed as follows:

$$\mathcal{L}_{aug} = \prod_{i=1}^N \prod_{j=1}^J \prod_{t=1}^T I_{\{\sum_{i=1}^N z_{ijt} = NS_{jt}\}} p_{ijt}^{z_{ijt}}, \quad (3)$$

where the indicator function ensures that the (augmented) individual choices are exactly consistent with the aggregate market shares. Using equation (3), the posterior density of the parameters and augmented data  $Z$  is proportional to the following expression:

$$f(Z, \theta, \bar{\theta}, D | S, X) \propto \left( \prod_{i=1}^N \prod_{t=1}^T \phi(\theta_{it}; \bar{\theta}, D) \prod_{j=1}^J I_{\{\sum_{i=1}^N z_{ijt} = NS_{jt}\}} p_{ijt}^{z_{ijt}} \right) \pi(\bar{\theta}, D). \quad (4)$$

where  $\theta$  is a matrix containing each of the vectors of individual coefficients ( $\theta_{it}$ );  $S$  denotes the observed data matrix with the market shares of each of the  $J$  alternatives in each period;  $X$  corresponds to a matrix containing marketing information (e.g., prices and brand dummies) for each of the  $J$  alternatives in each of  $T$  periods;  $\phi(\cdot; \bar{\theta}, D)$  is the density of a multivariate normal distribution with mean  $\bar{\theta}$  and variance-covariance matrix  $D$ ; and  $\pi(\bar{\theta}, D)$  is the hyperprior for  $\bar{\theta}$  and  $D$ , which is specified here as a standard Normal-Inverted Wishart prior (see Gelman *et al.*, 1995, p. 80); albeit, the results provided here can be adapted to other priors.

As previously mentioned, the estimation is implemented via Gibbs sampling by simulating each of the unknown parameters ( $\theta, \bar{\theta}, D$ ) from their full-conditional posterior distributions. We first note that the updating of  $\bar{\theta}$  and  $D$  from their full-conditional distributions (i.e., conditioning on  $\theta_{it}$ ) can be performed using standard conjugate methods (Allenby and Rossi, 2003). In addition, for this particular system of demand, as mentioned, the values of  $z_{ijt}$  remain fixed at all iterations in order to prevent the Markov Chain from experiencing label-switching problems. Therefore, the discussion here focuses only on the updating of  $\theta_{it}$ , the remaining parameters. However, in §3 for the panel case the sampling is more complex, since we also need to generate samples of  $Z$  from its posterior distribution.

The full-conditional posterior density of  $\theta_{it}$ , for the case of the first system of demand (inde-

pendent samples), is proportional to the following expression:

$$f(\theta_{it}|\ast) \propto \phi(\theta_{it}; \bar{\theta}, D) \prod_{j=1}^J p_{ijt}^{z_{ijt}}. \quad (5)$$

Since this density is from a non-conjugate form, we use a MH step to obtain posterior samples for  $\theta_{it}$ . One possibility is to use a symmetric random walk with a multivariate normal jumping kernel centered at the current value of  $\theta_{it}$  (Chib and Greenberg 1995). However, our numerical results suggest that, in terms of speed of convergence, this is not an efficient way to generate draws of  $\theta_{it}$ . Instead, we note from (5) that the full-conditional posterior distribution of  $\theta_{it}$  is equal to its multivariate normal prior density perturbed only by the likelihood of a single choice, as each consumer appears in one and only one time period. Therefore, a reasonable and simple approach to generate candidate draws ( $\theta_{it}^*$ ) is to sample these vectors from a proposal distribution equal to the normal prior (i.e.,  $\theta_{it}^* \sim N(\bar{\theta}^{(k)}, D^{(k)})$ ) updated at each iteration  $k$ . Therefore, in a given iteration  $k$  each of these newly drawn candidate vectors  $\theta_{it}^*$  should then be accepted according to the following MH probability:

$$\alpha_{\text{MH}, \theta_{it}} = \text{Prob}(\theta_{it}^{(k+1)} = \theta_{it}^*) = \min \left( \frac{\prod_{j=1}^J p_{ijt}^* z_{ijt}}{\prod_{j=1}^J p_{ijt}^{(k) z_{ijt}}}, 1 \right), \quad (6)$$

where  $p_{ijt}^*$  is the multinomial logit probability that the  $i^{\text{th}}$  consumer in the random sample from period  $t$  chooses alternative  $j$  when its coefficient  $\theta_{it}$  is equal to  $\theta_{it}^*$ ; and  $p_{ijt}^{(k)}$  is the corresponding probability when  $\theta_{it}$  is equal to  $\theta_{it}^{(k)}$ . Finally, we note that this particular choice of the proposal distribution corresponds to the third method described in Chib and Greenberg (1995, p. 330). In the next subsection we present the results from a numerical experiment demonstrating the efficacy of this approach. Given that other proposal distributions could also be considered, future research should evaluate the performance of alternative approaches for this estimation problem under different scenarios.

## 2.4 Numerical Experiments

Numerical simulation examples were conducted in order to illustrate the performance of the proposed method. The specific values of  $J$  (the number of brands),  $T$  (the number of time periods) and the distribution of  $\theta_{it}$  (i.e., the values of  $\bar{\theta}$  and  $D$ ) were chosen both to mimic extant research (Chen and Yang, 2007) and to provide realistic values on the probability scale without loss of generality. In each of these examples we consider  $J = 3$  brands,  $T = 50$  periods and  $N = 250$  consumers in each period. The utility function of each of these consumers includes three explanatory variables described below. The true mean of the individual coefficients ( $\theta_{it}$ ) was chosen as  $\bar{\theta} = (1 \ 1 \ -1)'$ . We specify three different values for the variance-covariance matrix ( $D$ ) to assess the robustness of the proposed method under varying degrees of heterogeneity and correlation in consumer tastes. The first corresponds to:

$$D^{(1)} = \begin{bmatrix} 3.0 & 0.5 & -0.5 \\ 0.5 & 1.0 & 0.0 \\ -0.5 & 0.0 & 2.0 \end{bmatrix}$$

while the remaining two correspond to  $D^{(2)} = I_3$  and  $D^{(3)} = 3 I_3$ , where  $I_3$  denotes the identity matrix with three rows and columns. For ease of exposition, we refer to these three scenarios as the correlation, low heterogeneity and high heterogeneity cases, respectively. In terms of the explanatory variables in the utility function, the first two correspond to intercepts for each of the first two brands, while the third was generated from a normal distribution  $N(0, \sigma_{x_3}^2)$ . For the first two cases (i.e., for  $D = D^{(1)}$  and  $D = D^{(2)}$ ),  $\sigma_{x_3} = 1$ . For the third case,  $\sigma_{x_3}$  equals 0.1. This smaller standard deviation is set in order to prevent the simulated choice probabilities from reaching values too close to either 0 or 1 (beyond 0.00001 and 0.99999).

The starting value for MCMC computation of  $\bar{\theta}$  for each of the three cases is  $\bar{\theta} = (0 \ 0 \ 0)'$ . The starting value for  $D$  corresponds to  $D = 0.1 I_3$  for the low heterogeneity case, while it is equal to  $D = I_3$  for the high heterogeneity and correlation cases. In addition, the following hyperprior distributions are specified:  $\bar{\theta} \sim N(0_3, 10^5 I_3)$  and  $D \sim \text{Inverted Wishart}(5 I_3, 5)$ , extremely weak hyperpriors; where  $0_3$  denotes a column vector with all of its three elements equal to zero. The

results for the correlation case are presented in the first block (independent samples, full sample) of Table 1 and they are based on a single run of 200,000 iterations where the last 100,000 are used for the estimation<sup>3</sup>.

== Insert Table 1 here ==

From the results in Table 1 we observe that the true values of  $\bar{\theta}$  and  $D$  are covered by their 95% posterior-probability intervals. In addition, we note that for this numerical example the proposal approach based on the prior  $\theta_{it}^* \sim N(\bar{\theta}, D)$  exhibited an acceptance rate between 52% and 63%. Therefore, although it does not achieve a perfect approximation of the full-conditional posterior density, it does not reject the candidate vectors of individual coefficients too often and appears to be mixing quite well. In summary, these results provide numerical support for the methods introduced in this section for the independent samples case.

## 2.5 Implementation issues

Given that the per-iteration times and memory requirements increase with the number of consumers in each period ( $N$ ), this creates a dimensionality problem<sup>4</sup>. In this section we propose two solutions to break this dimensionality problem. The first solution is based on subsampling. The idea is to randomly select (without replacement) the choices of a smaller number of consumers  $R < N$  prior to running the Gibbs sampler and then, based on this subsample of choices, estimate the posterior distribution of the parameters of the model<sup>5</sup>. Therefore, for each of the  $T$  random samples we randomly pick the choices of  $R$  consumers, where each consumer has the same probability of being selected. Using the same simulated data sets described in §2.4, we implemented this subsampling approach using  $R = 50$ . The results for the correlation case are reported in the second block (independent samples, subsampling) of Table 1, while those for the low and high

---

<sup>3</sup>Results for the remaining two cases (low and high heterogeneity) are available in Tables A1 and A2 in the technical appendix accompanying this manuscript which is available on the first author’s website: <http://faculty.fuqua.duke.edu/~amusalem/bio/index.htm>.

<sup>4</sup>Specifically, the values of  $N \cdot T$  vectors of coefficients must be stored and the same number of MH steps are necessary in order to update the coefficients of all consumers.

<sup>5</sup>For example, suppose  $N=1000$ ,  $R=100$ ,  $J = 2$  and that in a given period 700 consumers chose brand A and 300 consumers chose brand B. Then we implement subsampling by first constructing a fictitious data set  $Z$  consistent with the aggregate market share information and then we select at random and without replacement 100 of these 1000 choices in  $Z$ , where each of these 1000 choices has the same probability of being selected.

heterogeneity cases are presented in Tables A1 and A2 in the technical appendix. We observe that the true values of the parameters are again contained within their 95% posterior-probability intervals. As expected, the posterior standard deviations are higher than in the case where we use all the choices, i.e. when  $R = N$  (see first block in Table 1). In terms of computational speed, by using a subsample of  $R = 50$  choices per period instead of using  $R = 250$ , the per-iteration times are reduced from 0.021 to 0.004 seconds, expectedly a factor approximately equal to 5. These estimates were obtained using GAUSS on a 3.2 GHz Pentium 4 processor with 1 GB of RAM.

In summary, we show that it is possible to overcome the dimensionality problem by using an estimation procedure based on the choices of a smaller set of  $R < N$  consumers. This will be particularly important for empirical applications when a researcher needs to estimate consumer preferences using aggregate data that were generated by a large sample of consumers.

### 3 Consumer Panel Case

Assume that there are  $N$  consumers in the market, where  $N$  is known. In each period, each of these  $N$  consumers maximizes his/her latent utility by choosing the product with the highest utility. Denoting by  $y_{it}$  the choice of consumer  $i$  in period  $t$ , then:

$$\begin{aligned}
 y_{it} &= \operatorname{argmax}_j U_{ijt} \\
 &= \operatorname{argmax}_j V_{ijt} + \epsilon_{ijt} \\
 &= \operatorname{argmax}_j \theta'_i x_{jt} + \xi_{jt} + \epsilon_{ijt}.
 \end{aligned} \tag{7}$$

Note that the notation for this system has a similar, but not completely equivalent interpretation when compared to the definitions introduced in the previous section for the independent samples case. Specifically,  $U_{ijt}$  is the utility of alternative  $j$  for consumer  $i$  in period  $t$ ;  $V_{ijt}$  is defined as the sum of  $\theta'_i x_{jt}$  and  $\xi_{jt}$ ;  $\theta_i$  is a vector of coefficients for consumer  $i$ ; and,  $\epsilon_{ijt}$  is an individual-specific demand shock for the utility of alternative  $j$  for consumer  $i$  in week  $t$ . As before, we assume that  $\epsilon_{ijt}$  is distributed according to an Extreme Value (0,1) distribution and



we use a similar parametric assumption for the distribution of consumer coefficients by specifying that each  $\theta_i$  is i.i.d. Multivariate Normal with mean  $\bar{\theta}$  and variance-covariance matrix  $D$ . As we previously mentioned, we also assume that  $\xi_{jt} = 0$ , a condition that will be relaxed in §4. In summary, the main difference with the model in the previous section (independent samples) is that instead of having  $N \cdot T$  consumers where each of them makes only one purchase decision, we have  $N$  consumers making  $T$  purchase decisions each. Note that in this case we denote the corresponding consumer coefficients by  $\theta_i$  instead of  $\theta_{it}$ .

Lacking observed information on individual choices ( $y_{it}$ ), a data augmentation approach (Tanner and Wong, 1987; Chen and Yang, 2007) can be implemented by simulating the sequences of unobserved individual choices of these  $N$  consumers. These simulated sequences of choices must be consistent with the observed market shares (an *atypical* computational problem) and, consequently, they must satisfy the following hard constraint:

$$\sum_{i=1}^N z_{ijt} = NS_{jt}, \quad (8)$$

where, as in the previous section,  $z_{ijt}$  takes a value of 1 if consumer  $i$  chooses product  $j$  in period  $t$  (i.e., if  $y_{it} = j$ ), and is equal to 0, otherwise; and  $S_{jt}$  is the (observed) aggregate market share for brand  $j$  in period  $t$ . According to these definitions, the *augmented* likelihood can be written as follows:

$$\mathcal{L}_{aug} = \prod_{i=1}^N \prod_{j=1}^J \prod_{t=1}^T I_{\{\sum_{i=1}^N z_{ijt} = S_{jt}N\}} p_{ijt}^{z_{ijt}} \quad (9)$$

Note that in general there exist many values for  $Z$  (i.e., sequences of individual choices) that satisfy the market share constraint. However, different sequences have different implications for the pattern of consumer preferences and each of these sequences might achieve a different value of the likelihood and the posterior density of the parameters of the model. Therefore, it is necessary to estimate not only the posterior distribution of the individual coefficients  $\theta_i$  but also the corresponding distribution of the sequence of individual choices,  $Z$ . Using equation (9), the

posterior density of all the model parameters is proportional to the following expression:

$$f(Z, \theta, \bar{\theta}, D | S, X) \propto \left( \prod_{i=1}^N \phi(\theta_i; \bar{\theta}, D) \prod_{j=1}^J \prod_{t=1}^T I_{\{\sum_{i=1}^N z_{ijt} = S_{jt} N\}} p_{ijt}^{z_{ijt}} \right) \pi(\bar{\theta}, D) \quad (10)$$

where  $\theta$  is a matrix containing each of the vectors of individual coefficients ( $\theta_i$ ).

### 3.1 Estimation

The Gibbs sampling estimation method presented in this subsection relies on the fact, as in Section 2, that after conditioning on the current values of the individual choices,  $Z$ , the parameters  $\{\theta_i\}_{i=1}^R$ ,  $\bar{\theta}$  and  $D$  can be sampled using standard Bayesian methods (Allenby and Rossi, 2003) and hence are not described here. Therefore, we focus on the problem of how to generate draws of the augmented individual choices,  $Z$ . The approach introduced here is new to the literature and is a contribution of this research.

Under our method, the augmented individual choices,  $z_t$ , are drawn *directly* from their full-conditional posterior distribution (Gibbs sampling) instead of performing a MH step as in Chen and Yang (2007)<sup>6</sup>. As before, the sets of individual choices must satisfy the market share constraint defined in equation (8). This condition makes the draws of all individual choices interdependent. Consequently, drawing *all* these components directly and *simultaneously* would require the computation of the full-conditional posterior probability of *all* possible values of  $z_t$  that are consistent with the aggregate market shares. From a computational point of view, this is definitely a difficult task (i.e., infeasible) that would require a considerable amount of computing time to be accomplished. However, this computational problem can be circumvented if, instead of drawing the full set of individual choices for a given period ( $z_t$ ) at once, we draw subsets of individual choices that are space-filling (Liu, 1994). In particular, by considering pairs of choices, the full-conditional posterior distribution of each pair can be easily computed. Suppose that we consider the choices of consumers  $i_1$  and  $i_2$  in period  $t$ , while conditioning on all other parameters including the choices of all other consumers in period  $t$ . Then, using (10), the full-conditional posterior distribution of

---

<sup>6</sup>We provide more details about this alternative method in §3.3.

the choices of these two consumers in period  $t$  corresponds to:

$$f(z_{i_1t}, z_{i_2t} | *) = K \cdot I_{\{\sum_{i=1}^N z_{ijt} = S_{jt} N\}} \prod_{j=1}^J p_{i_1jt}^{z_{i_1jt}} p_{i_2jt}^{z_{i_2jt}} \quad (11)$$

where  $K$  is a normalization constant that depends on the values of all other parameters. Assuming that in a given iteration of the Markov Chain, the values in  $z_t$  satisfy the market share constraint, it is easy to verify that when the choices of all other consumers are held constant, there are only two instances of  $\{z_{i_1t}, z_{i_2t}\}$  that have a non-zero probability. The first corresponds to the current values of  $\{z_{i_1t}, z_{i_2t}\}$ , while in the second instance consumers  $i_1$  and  $i_2$  interchange their choices in period  $t$ . Note that any other configuration would violate the market share constraint. In addition, if the choices of both consumers in period  $t$  are the same, then only the current configuration of choices has a non-zero probability. Accordingly, the full-conditional posterior probability of the event where the choices of these two consumers take their current values can be computed as follows:

$$f(z_{i_1t}, z_{i_2t} | *) = \frac{\prod_{j=1}^J p_{i_1jt}^{z_{i_1jt}} p_{i_2jt}^{z_{i_2jt}}}{\prod_{j=1}^J p_{i_1jt}^{z_{i_1jt}} p_{i_2jt}^{z_{i_2jt}} + \prod_{j=1}^J p_{i_2jt}^{z_{i_2jt}} p_{i_1jt}^{z_{i_1jt}}} \quad (12)$$

Therefore,  $\{z_{i_1t}, z_{i_2t}\}$  can be drawn directly from its full-conditional posterior distribution by letting these choices remain at their current values with the probability specified in (12) or, otherwise, interchanging the choices of the two consumers. Using this result, the procedure for updating  $Z$  using a direct Gibbs step can be formalized as follows:

1. In each iteration ( $k$ ) and for each period ( $t$ ) randomly select  $N/2$  pairs of consumers without replacement and enumerate these pairs<sup>7</sup>. Let  $(i_{1pt}, i_{2pt})$  be the indices of consumers in pair  $p$  for period  $t$  and  $z_{i_{1pt}}^{(k)}$  and  $z_{i_{2pt}}^{(k)}$  their choices in that period in the current iteration  $k$ .
2. Starting from the first pair, successively and jointly draw  $z_{i_{1pt}}^{(k+1)}$  and  $z_{i_{2pt}}^{(k+1)}$  from their full-

---

<sup>7</sup>Note that this does not need to be necessarily done in pairs. However, the simplicity with which this step can be done changes dramatically when the number of components is increased (i.e., when using triplets, quadruplets, etc.). From a computational point of view, determining the optimal number of components that each subvector should have is certainly an important question for future research.

conditional posterior distributions according to the following steps:

- (a) If in the current iteration  $(k)$   $z_{i_{1pt}}^{(k)} = z_{i_{2pt}}^{(k)}$ , then, in order to satisfy the market share constraint (equation 8), these individual choices must stay at their current values. Therefore, assign  $z_{i_{1pt}}^{(k+1)} = z_{i_{1pt}}^{(k)}$  and  $z_{i_{2pt}}^{(k+1)} = z_{i_{2pt}}^{(k)}$ .
- (b) If in the current iteration  $(k)$   $z_{i_{1pt}}^{(k)} \neq z_{i_{2pt}}^{(k)}$ , then assign  $z_{i_{1pt}}^{(k+1)} = z_{i_{1pt}}^{(k)}$  and  $z_{i_{2pt}}^{(k+1)} = z_{i_{2pt}}^{(k)}$ , with the following probability:

$$f(z_{i_{1pt}}^{(k+1)} = z_{i_{1pt}}^{(k)}, z_{i_{2pt}}^{(k+1)} = z_{i_{2pt}}^{(k)} | *) = \frac{\prod_{j=1}^J p_{i_{1pt}jt}^{z_{i_{1pt}jt}^{(k)}} p_{i_{2pt}jt}^{z_{i_{2pt}jt}^{(k)}}}{\prod_{j=1}^J p_{i_{1pt}jt}^{z_{i_{1pt}jt}^{(k)}} p_{i_{2pt}jt}^{z_{i_{2pt}jt}^{(k)}} + \prod_{j=1}^J p_{i_{1pt}jt}^{z_{i_{2pt}jt}^{(k)}} p_{i_{2pt}jt}^{z_{i_{1pt}jt}^{(k)}}} \quad (13)$$

otherwise, exchange the choices of these two consumers assigning:  $z_{i_{1pt}}^{(k+1)} = z_{i_{2pt}}^{(k)}$  and  $z_{i_{2pt}}^{(k+1)} = z_{i_{1pt}}^{(k)}$ .

It is important to mention that the random pairing of latent individual choices (step 1) is necessary in order to guarantee that in every iteration any feasible set of individual choices has a positive probability of being sampled after some finite number of iterations<sup>8</sup>. Consequently, the Markov Chain in this Gibbs sampling method is irreducible - a necessary condition for the convergence of the Markov Chain to the posterior distribution of the parameters (Chib and Greenberg, 1995).

### 3.2 Numerical Experiments

We construct three numerical examples based on the same parameter values used in §2.4. We consider  $N = 250$  consumers,  $J = 3$  brands and  $T = 50$  periods. As before, the utility function of each of the simulated consumers includes intercepts for each of the first two brands, and a third variable generated from a Normal distribution. The mean of the individual coefficients corresponds to  $\bar{\theta}^{true} = (1 \ 1 \ -1)'$  and we use the same three values for D as in §2.4.

---

<sup>8</sup>Note that this would not be true if two components of  $y_t$  are always grouped together in all iterations.

For the MCMC estimation, we use the same starting values and hyperprior distributions for  $\bar{\theta}$  and  $D$  as in the previous section. In addition, the initial values of the individual choices were selected by randomly drawing  $z_t$  from a distribution that assigns the same probability to each set of choices satisfying the market share constraint. Using the direct-Gibbs method introduced in this paper, we obtained the results presented in the third block of Table 1 for the correlation case, while those for the low and high heterogeneity cases are presented in the third block of Tables A1 and A2 in the technical appendix. These results are based on a single run of 200,000 iterations with the last 100,000 used for estimation.

== Insert Table 2 here ==

From the results we observe that the true values of  $\bar{\theta}$  and  $D$  are covered by their 95% posterior-probability intervals. Comparing these results with those obtained in the case of independent samples, we notice that, in general, the posterior standard deviations for  $\bar{\theta}$  and  $D$  are of a similar magnitude. Finally, in terms of computational speed, the estimated per-iteration times are approximately 0.1 seconds. Note that this per-iteration time is higher than the one for the independent samples case (0.02 seconds), and the difference is primarily explained by the additional computational time required in the consumer panel case to update the sequences of individual choices ( $Z$ ).

### 3.3 Alternative Bayesian Methods

An alternative approach was proposed by Chen and Yang (2007). They assume that a market with  $M$  consumers (or clusters) with preference coefficients  $\theta_i \sim MVN(\bar{\theta}, D)$  can be approximated by considering  $R < M$  representative consumers with preference coefficients that are also distributed according to a  $MVN(\bar{\theta}, D)$  and  $M - R$  consumers with choice probabilities equal to the average of the choice probabilities of the  $R$  representative consumers (or clusters). Under this approach, it is necessary to augment the choices of the  $R$  representative consumers and this is accomplished by using a MH step.

In the context of our approach, it is also possible to simulate the unobserved choices using a MH step. For example, and as proposed in an earlier version of the Chen and Yang (2007)

article, one could use a MH step to update the set of individual choices corresponding to each period  $t$  ( $z_t$ ). This can be implemented by means of an independence chain, where at a given iteration  $k$  of the Markov Chain and for each period  $t$ , each set of  $N$  individual choices satisfying the market share constraint associated with period  $t$  (equation 8) has the same probability of being selected as the candidate set of choices ( $z_t^*$ ). The candidate  $z_t^*$  is then accepted according to a MH probability given by:

$$\alpha_{MH,z} = \text{Prob}(z_t^{(k+1)} = z_t^*) = \min \left( \frac{f(z_t^* | S_t, \{\theta_i\}_{i=1}^N, x_t)}{f(z_t^{(k)} | S_t, \{\theta_i\}_{i=1}^N, x_t)}, 1 \right), \quad (14)$$

and, otherwise,  $z_t^{(k+1)} = z_t^k$ , where  $f(\cdot | S_t, \{\theta_i\}_{i=1}^N, x_t)$  denotes the full-conditional posterior probability of  $z_t$ . A critical feature of this method is that all the individual choices in a given period are jointly sampled at each MH step. This has certain advantages and disadvantages. On the one hand, this ensures that a global criterion is used to accept or reject a candidate set of choices ( $z_t^*$ ). On the other hand, some components of a candidate set might improve the posterior distribution while others might reduce it. Since either all elements of  $z_t^*$  are accepted or all of them are rejected, it is not possible in the context of this algorithm to assign different acceptance probabilities to different subcomponents of  $z_t$ . Therefore, the joint sampling of all components in a given step induces some degree of inflexibility. This explains why the acceptance rate of candidate sets could reach levels well below 1%.

The acceptance rate can be improved, however, if smaller sets of individual choices are jointly sampled in blocks, instead of all at once. Specifically, in each iteration we can randomly divide (partition) the set of choices  $z_t$  in blocks of size  $B$ . Each of these blocks includes the choices of  $B$  consumers in period  $t$ . These choices are then updated using a MH step, where the candidate block of choices is generated from a distribution that assigns the same probability to every subset of  $B$  choices satisfying the market share constraint.

Using this modified version of the independence chain MH method and setting  $B = 10$  (i.e., performing  $250/10 = 25$  separate blocked MH steps instead of a single step) we obtained the set of results presented in the fourth block of Table 1 for the correlation case, and also in the fourth block

of Tables A1 and A2 for the low and high heterogeneity cases, respectively. From these results, we observe that the estimated posterior means and standard deviations for both the Modified MH and Direct-Gibbs methods are remarkably similar. In addition, using this modified method where the MH step is implemented in groups of  $B = 10$  components, the acceptance rate is approximately 15%, which is substantially higher than the corresponding one when  $B = N = 250$  (results available upon request). Consequently, we suggest that if a MH step is utilized for augmenting the sequences of latent choices of a panel of consumers, it might be more efficient to break the independence chain into sub-blocks of size  $B$ , where the optimal value for  $B$  is an area for future research, which is conceptually similar to the problem studied by Gelman *et al.* (1996).

### 3.4 Implementation issues

For each of the methods discussed in §3.1 and §3.3, the per-iteration times and memory requirements increase with  $N$ . In this respect, we acknowledge that there are instances in which simulating the unobserved choices of  $N$  consumers might be infeasible from a computational point of view. For this reason, an alternative solution to this dimensionality problem will be discussed in Section 4. Specifically, we show that there is an asymptotic link between the two systems of demand discussed in this paper: panel of consumers and independent random samples. Based on this link, we propose the use of the estimation methods developed for the first system, which are better suited for handling large samples of consumers (because augmentation of  $Z$  is not needed and subsampling solutions are available), for the estimation of the second system of demand. As we will discuss in the next section, this result is not only relevant for the estimation of the demand of a panel of consumers, but also for the corresponding one of a hybrid system where each consumer makes purchase decisions in some of the  $T$  periods.

## 4 Asymptotic equivalence of both systems of demand

In both systems of demand, the dependent variable corresponds to the aggregate market share of each alternative for each period. Since this quantity can be written as an average of i.i.d. random

variables (i.e,  $S_{jt} = \frac{1}{N} \sum_{i=1}^N z_{ijt}$ ), it is possible to apply the Weak Law of Large Numbers (WLLN) for this class of random variables.

**Proposition 4.1:** *Assuming the same values of  $\bar{\theta}$ ,  $D$  and  $\xi_t$  for both systems of demand (independent samples and consumer panel), the observed aggregate data ( $S_{jt}$ ) from these two systems converges in probability to the same values as  $N \rightarrow \infty$ .*

**Proof:** Applying the WLLN for i.i.d. random variables to both systems of demand, it is verified that  $S_{jt}$  converges in probability to  $E(z_{ijt}|\bar{\theta}, D, \{\xi_{jt}\}_{j=1}^J)$ . This expectation can be computed in both cases as:

$$E(z_{ijt}|\bar{\theta}, D, \{\xi_{jt}\}_{j=1}^J) = \int \frac{e^{\theta' x_{jt} + \xi_{jt}}}{\sum_{k=1}^J e^{\theta' x_{kt} + \xi_{kt}}} \phi(\theta; \bar{\theta}, D) d\theta.$$

It is important to mention that this property can be easily generalized to other demand models with different assumptions about the heterogeneity of consumer preferences (e.g., finite mixture instead of multivariate normal) and the distribution of the individual-specific demand shocks  $\epsilon_{ijt}$  (e.g., multivariate normal instead of extreme value). Similarly, it is easy to show that the two systems are also asymptotically equivalent to a hybrid system in which consumers make purchase decisions in some but not all of the  $T$  periods (as long as the probability of a consumer making a purchase decision in a given period is independent of her vector of preferences).

In order to provide some numerical evidence of the convergence of the panel and independent samples systems of demand we performed the following simulation experiment, where  $\bar{\theta} = (1 \ 1 \ -1)$  and  $D = I_3$ . Specifically, we simulated the market shares of three brands for each of the two systems of demand using a value of  $N = 100,000$  choices per period. We also included common demand shocks in the utility function of each consumer, where each vector of common demand shocks for a given period ( $\xi_t$ ) is i.i.d. multivariate normal with zero mean and variance covariance matrix equal to the identity matrix (i.e.,  $\Sigma_\xi = I_3$ ). For each of the  $T = 50$  periods we used the same values of the explanatory variables ( $x_{jt}$ ) and common demand shocks ( $\xi_{jt}$ ) for both systems. From our results, we noted a very small and insignificant difference in the aggregate behavior of these two systems given that the largest discrepancy in market shares between these two demand system was smaller than 0.0083 for this numerical example (detailed results are available from the



authors upon request).

Proposition 4.1 has important implications for the estimation of the second system of demand. As  $N$  increases, the aggregate behavior of a consumer panel gets stochastically closer to the behavior of a system where independent random samples generate the aggregate data for each period. Therefore, for sufficiently large  $N$  we can treat the consumer panel data as if they were generated by independent random samples of consumers. Accordingly, we can use the solutions proposed in §2.5 to handle large samples of consumers. Moreover, it is also possible to show that the information obtained about the model parameters under both demand systems is also asymptotically equivalent when  $N$  approaches infinity. This result is formalized in the next theorem<sup>9</sup>.

**Theorem 4.1:** *Assume that the same hyperprior distribution,  $\pi(\bar{\theta}, D, \Sigma)$ , is used in the estimation of the independent samples and consumer panel demand systems. When an infinite number of consumers make purchase decisions in each period, the posterior distributions of the parameters of the model  $f(\bar{\theta}, D, \Sigma \mid S, X)$  under both demand systems (independent samples and panel of consumers) are equal.*

**Proof:** Please refer to the Appendix.

Finally, to illustrate the implications of this asymptotic equivalence, we generated data for a *panel of consumers*, where consumers choose between three alternatives plus an outside good in each of  $T = 50$  time periods. The utility function of each of the three alternatives includes three brand dummies and a fourth variable generated from a standard normal distribution. We generated common demand shocks for each of the three brands. In empirical applications these common demand shocks may be correlated with some of the variables in the utility function (e.g., prices). Consequently, in this simulation experiment we allow these shocks to be correlated with the values of the fourth explanatory variable ( $x_{jt,4}$ ). Specifically, we assume  $x_{jt,4} = w'_{jt}\delta_j + \eta_{jt}$ , where  $w_{jt}$  is a vector of observed instruments for  $x_{t,4}$ ,  $\delta_j$  is a vector of coefficients and  $\eta_{jt}$  is an error term which is independent of the instruments, but can be potentially correlated with the demand shocks. In particular, each vector  $w_{jt}$  has two components, where the first component

---

<sup>9</sup>We thank Peter Rossi for suggesting to us the importance of proving this theoretical result.

equals one (in order to estimate an intercept for  $x_{jt,4}$ ) and the second component is generated from a uniform distribution between 0 and 1. In terms of  $\delta$ ,  $\delta_1 = (0.5, 1.0)'$ ,  $\delta_2 = (0.5, 2.0)'$  and  $\delta_3 = (0.5, 0.5)'$ . Finally, each vector  $(\eta_{1t}, \eta_{2t}, \eta_{3t}, \xi_{1t}, \xi_{2t}, \xi_{3t})'$  is assumed to be i.i.d. multivariate normal with zero mean and variance-covariance matrix:

$$\Sigma = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.3 & 0.2 & -0.2 \\ 0.0 & 1.0 & 0.0 & -0.2 & 0.3 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.2 & 0.3 \\ 0.3 & -0.2 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.2 & 0.3 & 0.2 & 0.0 & 0.5 & 0.0 \\ -0.2 & 0.0 & 0.3 & 0.0 & 0.0 & 0.5 \end{bmatrix}$$

The estimation is implemented by treating the panel data as if they were generated by independent samples and subsampling the choices of  $R = 250$  consumers in each period<sup>10</sup>. The hyperpriors correspond to:  $\bar{\theta} \sim N(0_4, 100 I_4)$ ,  $D \sim \text{Inverted Wishart}(6 I_4, 6)$ ,  $\Sigma \sim \text{Inverted Wishart}(I_6, 1)$  and  $\delta_j \sim N(0_2, 10^5 I_2)$ . The results are based on a single run of 400,000 iterations where the last 200,000 were used for the estimation of the posterior distribution of the parameters (see Table 2).

== Insert Table 2 here ==

According to these results, the estimated values are close to the truth and the true values are covered by their 95% posterior probability intervals (except for  $D_{23}$ ). We also note that the estimated posterior mean and median of each of the parameters that capture the correlation between  $\eta_t$  and the common demand shocks  $\xi_t$  (i.e., the non-zero off-diagonal terms of  $\Sigma$ ) have the right sign.

---

<sup>10</sup>The common demand shocks  $\xi_t$  can be sampled using a MH step with a Normal random walk (see Yang *et al.*, 2003). However, in the case of independent samples and when the utility function includes brand intercepts, it is possible to implement a more efficient procedure that eliminates the need of performing MH steps for each vector  $\xi_t$ . The results presented in this paper are based on this idea and represent a computational contribution of this research. Details about the estimation procedure are presented in Technical Appendix A.

## 5 Model Selection

A very important issue for applied researchers is to determine which model provides the best approximation of the underlying data generating process. From a Bayesian perspective, the marginal likelihood, i.e. the probability of the observed data given a certain model, provides a criterion for answering this question and, therefore, for implementing a model selection procedure. One approach, which has some limitations as explained by Newton and Raftery (1994), is to use a harmonic mean method. Accordingly, in the independent samples case, the marginal probability of the aggregate data ( $A$ ) given a certain model  $\mathcal{M}$  can be computed as follows:

$$\hat{p}(A|\mathcal{M}) = \frac{1}{\frac{1}{m} \sum_{l=1}^m \frac{1}{p(Z|\varphi^{(l)})}}, \quad (15)$$

where  $Z$  is the set of choice indicators for each of the  $N \cdot T$  purchase occasions;  $\varphi$  denotes all the parameters of the model;  $\varphi^{(l)}$  denotes the  $l^{\text{th}}$  draw of  $\varphi$  from its posterior distribution (which can be directly obtained from the MCMC output); and  $m$  denotes the total number of draws used in the estimation of the marginal density.

In the panel of consumers case, a very similar expression can be used. Specifically, it can be shown that the marginal probability can be estimated as follows:

$$\hat{p}(A|\mathcal{M}) = \frac{|\Omega|}{\frac{1}{m} \sum_{l=1}^m \frac{1}{p(Z^{(l)}|\varphi^{(l)})}}, \quad (16)$$

where  $|\Omega|$  denotes the number of different configurations of individual choices that are consistent with the aggregate data, i.e. the number of elements in the set  $\left\{ Z : \sum_{i=1}^N z_{ijt} = N_{jt}, \forall j, t \right\}$ ; and  $Z^{(l)}$  is the  $l^{\text{th}}$  draw of  $Z$  from its posterior distribution, which can be directly obtained from the MCMC output. In practice, when comparing different models it will often be the case that the quantity  $|\Omega|$  will be the same for all of them (for example, when comparing models that only differ in terms of the explanatory variables ( $x_t$ ) that are included in the utility function). Therefore, in many instances, this quantity may be ignored for the purposes of computing marginal likelihood

ratios. Finally, we note that if a model for the covariates is also specified, then the likelihood of the covariates must also be considered when estimating the marginal likelihood (Manchanda *et al.* 2004).

## 6 Empirical Application

In order to provide a richer demonstration of our approach in a real setting, we analyze data regarding the demand for facial tissue products. We use 104 weeks of data collected by Information Resources Inc. (IRI), a leading scanner-data company, for the period spanning 08/30/1992 to 08/21/1994. These data contain information on weekly volume sales, volume prices and feature and display activity for each UPC in the Buffalo InfoScan market. We use the first 52 weeks of data for calibration while the last 52 weeks are used for validation (hold-out sample). We focus on three named brands which, in addition to private label brands (PL), account for more than 95% of the total volume sales in this market. These three brands correspond to Kleenex (KL), Puffs (PF) and Scotties (SC). In this application we model the choices of consumers among these four alternatives and a no-purchase option using eight explanatory variables. The first four variables are brand dummies for KL, PL, PF and SC. Since the sales of facial tissue may be subject to seasonal fluctuations, we include an explanatory variable equal to the natural logarithm of the average monthly temperature in this market (*lntemp*) in the utility of each of the four brands<sup>11</sup>. The next two explanatory variables correspond to dummy variables for feature (*feat*) and display (*disp*). The last variable corresponds to volume price (*price*), which is measured in dollars per 20 units of facial tissue. Finally, in order to compute the share of the no-purchase option it was necessary to estimate the size of the potential market ( $N$ ). For this purpose, we used U.S. Census data<sup>12</sup> to estimate the population in this market and then we multiplied this number by the average

---

<sup>11</sup>These data were obtained from the “Climatological Data Annual Summary” for the state of New York which is compiled by the National Oceanic and Atmospheric Administration (NOAA). Since the mean temperature data are available on a monthly basis, while all other variables are on a weekly basis, we applied the linear filter proposed by Slade (1995) and also used by Besanko *et al.* (2003). Accordingly, we assign to  $\text{temp}_t$  the mean temperature of the corresponding month and then we smooth the series by defining:  $\text{temp}_t^s = 0.25\text{temp}_{t-1} + 0.5\text{temp}_t + 0.25\text{temp}_{t+1}$ . In addition, we also mean centered the natural logarithm of this smoothed variable to facilitate interpretation.

<sup>12</sup>Source: <http://www.census.gov/popest/archives/1990s> (accessed on 01/30/2006).

volume of facial tissue purchased by each buyer (32.7 units/week), which was obtained from the IRI Marketing Factbook. Summary statistics are reported in Table 3. In terms of the estimation procedure, we use the method described in §4 and in Technical Appendix A2 (with subsampling and setting  $R=250$ ). Following Nevo (2001) we utilized instrumental variables corresponding to volume prices for facial tissue in other InfoScan markets to control for price endogeneity effects. In addition, we specify the following hyperprior distributions:  $\bar{\theta} \sim N(0_8, 50 I_8)$  and  $D \sim IW((8+2) I_8, 8+2)$  and  $\Sigma \sim IW(0.01 I_8, -1)$ <sup>13</sup>.

== Insert Table 3 here ==

## 6.1 Results

In this subsection, we present a summary of results around a number of features of the model: brand intercepts, seasonality and price endogeneity. In §6.2 and §6.3 we discuss inferences derived from the model, but not directly from the estimated values of the demand parameters. This initial set of results is presented in Table 4. For space considerations, we only report results for the off-diagonal elements of  $D$  (third block) and  $\Sigma$  (fifth block) that are significantly different from zero at the 90% posterior-probability level. These results are based on three parallel chains of 1,200,000 iterations, where the last 600,000 were used for estimation. Note that a larger run was done in this case as we wanted to ensure convergence, an issue easily asserted in the previous simulations (where the truth is known), but not here<sup>14</sup>.

== Insert Table 4 here ==

From the results in the first block of Table 4, KL seems to be on average intrinsically more preferred than the other brands (the 95% posterior interval of  $\bar{\theta}_{KL}$  is strictly above 0), although there is substantial heterogeneity in terms of consumer tastes. In addition, the mean of the price coefficient is negative and its 95% posterior-probability interval is entirely contained in the

---

<sup>13</sup>We used a relatively stronger but still weakly informative prior distributions for  $\bar{\theta}$  in order to prevent the Markov Chain from moving towards extreme and unrealistic values.

<sup>14</sup>Convergence was assessed by computing the Gelman-Rubin statistic (Gelman and Rubin, 1992). The mean value of the Gelman-Rubin statistic across all scalar estimands was equal to 1.03, while the standard deviation was equal to 0.04 and the maximum value was equal to 1.21 (well within the criterion's reasonable values).

negative domain, a result of face validity. These results are consistent with the summary statistics reported in Table 3 where we observe that even though the average volume prices of KL and PF are similar, the market share of KL is 126% higher than the market share of PF. In addition, PL and SC appear to be compensating for their lower intrinsic preference relative to KL by charging substantially lower prices, which allow them to enjoy market shares similar to those from PF.

In terms of the temporal fluctuations of the demand, we observe that although the mean effect of the seasonality variable is not significantly different from zero, there is substantial variation of this effect across consumers. This can be interpreted as the coexistence of two segments of consumers, one of them that it is more likely to purchase facial tissue in the coldest months of the year (possibly due to the higher incidence of influenza and the common cold) and another one more likely to purchase in the warmest months (perhaps due to the higher incidence of allergies during these months of the year). Some interesting findings can also be obtained by analyzing the off-diagonal elements of  $D$ . In particular, given that  $D_{temp,price}$  and  $D_{feat,price}$  are estimated to be negative, consumers that are relatively more likely to purchase facial tissue in the warmest months of the year and consumers that respond more to feature advertisements are estimated to be more price sensitive.

In addition, the results for  $\Sigma$  suggest that the weekly variation in the mean utility of each brand is not fully captured by the explanatory variables included in the model. In particular, these fluctuations are higher in magnitude for the utility of SC for which the estimated posterior mean of  $\Sigma_{\xi_{sc},\xi_{sc}}$  is equal to 1.403. Moreover, the covariances between  $\eta$  and the common demand shocks  $\xi_t$  are not significant with the exception of  $\Sigma_{\eta_{kl},\xi_{sc}}$ , which is estimated to be significantly negative at a 90% level. In fact, the posterior mean of this covariance is equal to -0.238 and the corresponding posterior mean of the correlation between  $\eta_{kl}$  and  $\xi_{sc}$  is equal to -0.401. This provides evidence consistent with the hypothesis that prices are set by firms with some knowledge of these unobserved (to the researcher) demand shocks (Berry 1994).

## 6.2 Price Elasticity of Demand

From a managerial point of view, the set of own- and cross-price elasticities provide an interesting characterization of the competition among the four brands under study. These elasticities can be easily estimated using the MCMC output. In Table 5 we report the estimated elasticities for a period with average levels of *lntemp*, *feat*, *disp* and *price* and in which common demand shocks ( $\xi_{jt}$ ) are equal to zero. From these results, we note some important asymmetries in the pattern of substitution among brands. In particular, we note that the influence of the prices of brand (KL) on the sales of PL and SC is much stronger than the effect of the prices of those brands on the sales of KL. Finally, we also note that the magnitude of each of the posterior means of the own-price elasticities is greater than 1 (except for PF), which is consistent with profit-maximization behavior.

== Insert Table 5 here ==

## 6.3 Out-of-Sample Validation

In order to validate our empirical results presented here, we provide forecasts for the market shares of each of the brands, the private label and the outside good in the 52 week hold-out sample. The forecasted values correspond to the estimated posterior mean of the market shares of each of these brands in the hold-out periods. These results are presented in Figures 1a, 1b, 1c, 1d and 1e, where we also display the 2.5% and 97.5% posterior quantiles for these quantities. In general, we observe that for each of the brands the posterior means are close to the observed realizations of the demand. In terms of the accuracy of the forecasted values, the mean absolute percentage errors (MAPE) are 19.3%, 21.5%, 30.7%, 38.1% and 2.9%, for KL, PL, PF, SC and the outside good, respectively; while the mean absolute deviations (MAD) are 0.016, 0.005, 0.009, 0.009 and 0.023, respectively.

== Insert Figures 1a, 1b, 1c, 1d and 1e here ==

## 7 Conclusions

In this article, we discuss the problem of estimating the distribution of consumer preferences from aggregate market share data using Bayesian methods. We analyzed two different systems of demand and for each of them we introduced estimation methods and presented results from numerical simulations.

For the first system of demand (independent samples), we presented estimation methods that do not require the simulation of unobserved individual choices and are similar in terms of computational requirements to those used when disaggregate data are available. Moreover, these methods can be easily applied in situations where a very large number of consumers generate the aggregate data for each period. For the second system of demand (consumer panel), we used a data augmentation strategy by which we simulate the unobserved individual choices from their posterior distribution. A key component of the data augmentation methods described here is the mechanism that generates draws of the unobserved individual choices. In this respect, we show that by using the direct-Gibbs and the modified MH methods introduced in this paper we are able to recover the true values of the parameters of the demand model.

One possible extension in the case of the direct-Gibbs method is related to the number of components that are jointly sampled, which in this paper is implemented by pairs. Obviously, the same procedure could be implemented using groups of 3, 4 or more components. In the context of Gibbs-sampling, drawing all individual choices simultaneously is theoretically more efficient than drawing subvectors (Liu, 1994). However, the computational cost of this efficiency gain must also be considered. Therefore, determining the optimal number of components in each group is an important research question from an applied point of view.

The asymptotic link between the two systems of demand introduced in Section 4 provides a theoretical justification for treating the data generated by a consumer panel (or a hybrid system) as if they were generated by independent samples of consumers. This allows us to use methods that are computationally more efficient for the purposes of handling a large sample of consumers, as in the case of the empirical application discussed in §6. In terms of future work, we believe



that there are great opportunities for extending the methods introduced in this paper to handle other situations in which the researcher only observes aggregate data about other aspects of the behavior of consumers. Some examples may include coupon usage, advertisement exposure and out of stocks. In particular, Musalem *et al.* (2007) present an extension of these methods to estimate the distribution of preferences and coupon usage among consumers for a case in which only aggregate information about consumer choices and coupon redemption is available. Another application in progress corresponds to the estimation of demand models from aggregate data with incomplete information about product availability. Finally, the methodology could also be generalized to incorporate and test alternative supply models (e.g., Yang *et al.*, 2001; Romeo, 2007; Duan and Mela, 2007).

In summary, this article adds to a growing literature in an area that we believe deserves considerable attention. We hope that the new methods presented in this paper might become useful to researchers interested in the application of Bayesian methods in the study of marketing and economics problems based on aggregate or limited data.

Table 1  
 Results: Estimated posterior mean, standard deviation and quantiles for  $\bar{\theta}$  and  $D$  (correlation).

Demand System	Method	$R$ (ind. Samples), $B$ (panel)		$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$D_{11}$	$D_{22}$	$D_{33}$	$D_{12}$	$D_{13}$	$D_{23}$
Ind. Samples	Full Sample	250	mean	0.980	0.928	-0.932	1.644	0.892	1.742	0.265	-0.271	0.061
			std.dev.	0.097	0.110	0.090	1.413	0.429	0.400	0.444	0.249	0.171
			2.5%	0.826	0.775	-1.202	0.264	0.365	1.226	-0.304	-0.929	-0.263
			50.0%	0.964	0.899	-0.915	1.269	0.793	1.667	0.132	-0.236	0.060
			97.5%	1.201	1.197	-0.808	6.558	1.923	2.937	1.394	0.139	0.429
Ind. Samples	Subsampling	50	mean	0.908	0.957	-0.940	2.341	1.139	1.964	0.437	-0.031	0.097
			std.dev.	0.189	0.203	0.125	2.026	0.728	0.684	1.001	0.454	0.372
			2.5%	0.599	0.668	-1.262	0.535	0.365	1.057	-0.774	-0.972	-0.587
			50.0%	0.883	0.917	-0.921	1.710	0.940	1.838	0.197	-0.011	0.082
			97.5%	1.384	1.518	-0.750	8.784	3.326	3.788	3.359	0.862	0.894
Panel	Gibbs	2	mean	1.071	0.993	-0.977	2.100	1.204	1.659	0.789	-0.110	-0.149
			std.dev.	0.156	0.150	0.110	1.047	0.634	0.343	0.666	0.193	0.167
			2.5%	0.826	0.761	-1.210	0.616	0.412	1.112	-0.030	-0.510	-0.481
			50.0%	1.049	0.970	-0.970	1.918	1.053	1.616	0.636	-0.106	-0.151
			97.5%	1.437	1.350	-0.780	4.770	2.769	2.449	2.563	0.273	0.193
Panel	MH	10	mean	1.135	1.049	-1.022	2.573	1.673	1.863	1.092	-0.102	-0.155
			std.dev.	0.182	0.175	0.135	1.454	0.969	0.488	0.814	0.224	0.190
			2.5%	0.850	0.773	-1.327	0.625	0.492	1.190	-0.007	-0.574	-0.529
			50.0%	1.110	1.026	-1.007	2.239	1.414	1.760	0.929	-0.095	-0.156
			97.5%	1.548	1.436	-0.795	6.115	4.089	3.046	3.020	0.332	0.230
True Values				1.000	1.000	-1.000	3.000	1.000	2.000	0.500	-0.500	0.000

Table 2  
 Results: Estimated posterior mean, standard deviation and quantiles for  $\bar{\theta}$  and  $D$   
 (consumer panel data, endogeneity, subsampling R=250).

	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	$D_{11}$	$D_{22}$	$D_{33}$	$D_{44}$	$D_{12}$	$D_{13}$	$D_{14}$	$D_{23}$	$D_{24}$	$D_{34}$
mean	0.804	0.925	0.970	-1.006	1.453	1.113	1.849	1.082	-0.467	-0.461	-0.320	0.672	0.108	-0.244
std.dev.	0.137	0.142	0.147	0.084	0.579	0.377	0.866	0.190	0.308	0.464	0.198	0.440	0.196	0.239
2.5%	0.525	0.657	0.638	-1.176	0.588	0.572	0.646	0.744	-1.186	-1.358	-0.731	0.073	-0.260	-0.668
50.0%	0.809	0.922	0.975	-1.002	1.387	1.043	1.750	1.068	-0.431	-0.417	-0.313	0.593	0.108	-0.262
97.5%	1.068	1.209	1.248	-0.856	2.810	2.082	4.173	1.512	0.059	0.306	0.032	1.737	0.497	0.243
True Values	1.000	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000

	$\delta_{11}$	$\delta_{12}$	$\delta_{21}$	$\delta_{22}$	$\delta_{31}$	$\delta_{32}$	$\Sigma_{11}$	$\Sigma_{22}$	$\Sigma_{33}$	$\Sigma_{44}$	$\Sigma_{55}$	$\Sigma_{66}$	$\Sigma_{12}$	$\Sigma_{13}$
mean	0.441	1.034	0.488	2.306	0.663	0.430	1.206	0.924	1.214	0.587	0.445	0.722	0.083	0.000
std.dev.	0.201	0.331	0.194	0.291	0.203	0.321	0.178	0.138	0.180	0.128	0.102	0.158	0.110	0.128
2.5%	-0.043	0.393	0.101	1.734	0.259	-0.189	0.907	0.691	0.915	0.381	0.281	0.468	-0.131	-0.252
50.0%	0.441	1.032	0.489	2.302	0.665	0.427	1.189	0.910	1.197	0.572	0.432	0.703	0.082	0.000
97.5%	0.833	1.691	0.867	2.875	1.054	1.067	1.604	1.234	1.619	0.880	0.675	1.078	0.306	0.255
True Values	0.500	1.000	0.500	2.000	0.500	0.500	1.000	1.000	1.000	0.500	0.500	0.500	0.000	0.000

	$\Sigma_{14}$	$\Sigma_{15}$	$\Sigma_{16}$	$\Sigma_{23}$	$\Sigma_{24}$	$\Sigma_{25}$	$\Sigma_{26}$	$\Sigma_{34}$	$\Sigma_{35}$	$\Sigma_{36}$	$\Sigma_{45}$	$\Sigma_{46}$	$\Sigma_{56}$
mean	0.323	0.250	-0.242	-0.195	-0.272	0.210	-0.153	-0.044	0.313	0.384	-0.096	-0.141	-0.032
std.dev.	0.150	0.099	0.112	0.112	0.089	0.093	0.100	0.100	0.104	0.165	0.072	0.092	0.075
2.5%	-0.035	0.075	-0.476	-0.428	-0.464	0.040	-0.358	-0.242	0.127	0.080	-0.244	-0.341	-0.177
50.0%	0.321	0.243	-0.238	-0.191	-0.266	0.206	-0.151	-0.044	0.307	0.378	-0.095	-0.135	-0.034
97.5%	0.629	0.462	-0.032	0.013	-0.114	0.404	0.039	0.154	0.535	0.728	0.044	0.022	0.123
True Values	0.300	0.200	-0.200	0.000	-0.200	0.300	0.000	0.000	0.200	0.300	0.000	0.000	0.000

Table 3

Summary Statistics for market share and prices of the brands in the Facial Tissue data set.

Variable*	Mean	Std. Dev.	Min.	Max.	Observations
$S_{kl}$	0.068	0.054	0.033	0.311	52
$S_{pl}$	0.030	0.019	0.016	0.129	52
$S_{pf}$	0.030	0.015	0.017	0.113	52
$S_{sc}$	0.033	0.043	0.011	0.224	52
$feat_{kl}$	0.827	0.382	0.000	1.000	52
$feat_{pl}$	0.385	0.491	0.000	1.000	52
$feat_{pf}$	0.558	0.502	0.000	1.000	52
$feat_{sc}$	0.827	0.382	0.000	1.000	52
$disp_{kl}$	0.885	0.323	0.000	1.000	52
$disp_{pl}$	0.923	0.269	1.000	1.000	52
$disp_{pf}$	0.712	0.457	0.000	1.000	52
$disp_{sc}$	0.769	0.425	1.000	1.000	52
$price_{kl}$	3.907	0.605	2.067	4.538	52
$price_{pl}$	2.384	0.189	1.772	2.726	52
$price_{pf}$	3.949	0.271	2.923	4.388	52
$price_{sc}$	2.579	0.281	1.682	3.089	52
$temp$	47.318	17.192	17.200	73.400	52

\* KL: Kleenex, PL: Private Label, PF: Puffs, SC: Scotties.

Table 4

Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles for  $\bar{\theta}$ ,  $D$  and  $\Sigma$  (facial tissue data, subsampling, R=250).

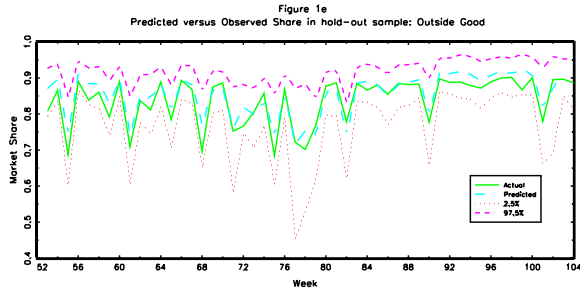
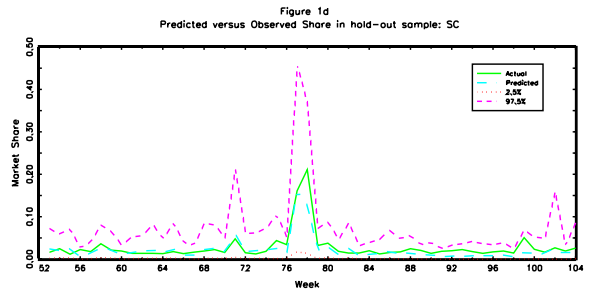
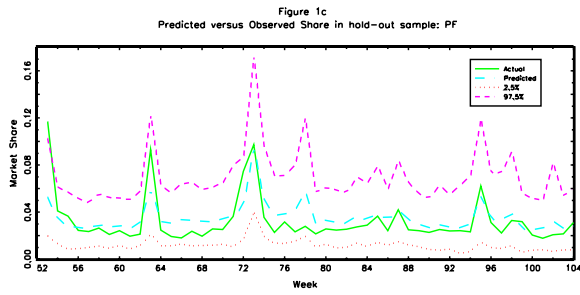
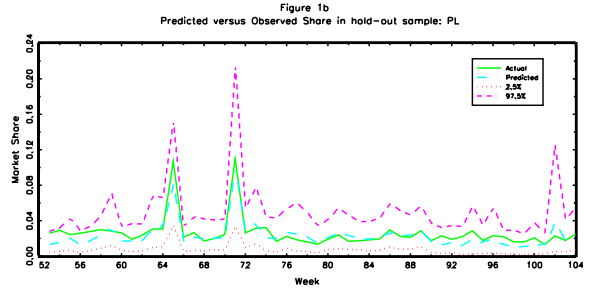
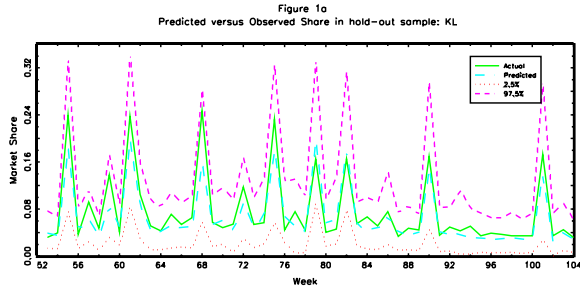
	mean	(s.d.)	2.5%	97.5%
$\bar{\theta}_{kl}$	3.402	(1.227)	1.139	6.254
$\bar{\theta}_{pl}$	-0.435	(1.212)	-3.071	1.754
$\bar{\theta}_{pf}$	0.988	(1.990)	-3.951	3.966
$\bar{\theta}_{sc}$	-0.044	(1.188)	-2.185	2.163
$\bar{\theta}_{intemp}$	0.104	(0.719)	-1.175	1.601
$\bar{\theta}_{feat}$	1.425	(1.017)	-0.182	3.802
$\bar{\theta}_{disp}$	2.053	(1.877)	-0.667	6.066
$\bar{\theta}_{price}$	-4.296	(0.736)	-5.911	-3.082
$D_{kl,kl}$	4.209	(3.108)	0.987	12.870
$D_{pl,pl}$	3.967	(2.894)	0.980	12.440
$D_{pf,pf}$	6.581	(7.239)	0.988	27.880
$D_{sc,sc}$	4.181	(2.810)	1.093	11.530
$D_{temp,temp}$	6.421	(4.185)	1.440	16.800
$D_{feat,feat}$	3.160	(2.129)	0.869	9.090
$D_{disp,disp}$	5.143	(4.834)	0.770	18.400
$D_{price,price}$	5.734	(2.747)	2.091	12.260
$D_{temp,price}$	-4.238	(2.975)	-11.320	-0.260
$D_{feat,price}$	-2.237	(1.789)	-6.846	0.207
$\Sigma_{\eta_{kl},\eta_{kl}}$	0.258	(0.068)	0.157	0.420
$\Sigma_{\eta_{pl},\eta_{pl}}$	0.032	(0.008)	0.020	0.052
$\Sigma_{\eta_{pf},\eta_{pf}}$	0.057	(0.014)	0.036	0.092
$\Sigma_{\eta_{sc},\eta_{sc}}$	0.058	(0.014)	0.036	0.092
$\Sigma_{\xi_{kl},\xi_{kl}}$	0.606	(0.463)	0.106	1.810
$\Sigma_{\xi_{pl},\xi_{pl}}$	0.354	(0.312)	0.056	1.213
$\Sigma_{\xi_{pf},\xi_{pf}}$	0.523	(0.453)	0.061	1.732
$\Sigma_{\xi_{sc},\xi_{sc}}$	1.403	(0.665)	0.564	3.108
$\Sigma_{\eta_{kl},\xi_{sc}}$	-0.238	(0.148)	-0.572	0.015
$\Sigma_{\xi_{pl},\xi_{sc}}$	0.560	(0.421)	0.079	1.672

Table 5

Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles of the own- and cross-price elasticities.

		KL	PL	PF	SC
mean	KL	-1.702	0.091	-0.056	0.092
	PL	0.350	-2.571	0.082	0.177
	PF	-0.092	0.044	-0.949	0.056
	SC	0.392	0.197	0.116	-3.294
s.d.	KL	0.694	0.082	0.242	0.080
	PL	0.299	0.846	0.172	0.120
	PF	0.464	0.089	0.937	0.084
	SC	0.322	0.130	0.174	1.000
2.5%	KL	-3.115	-0.053	-0.626	-0.033
	PL	-0.204	-4.221	-0.289	0.010
	PF	-1.151	-0.137	-2.897	-0.072
	SC	-0.160	0.012	-0.183	-5.285
97.5%	KL	-0.380	0.266	0.333	0.274
	PL	0.973	-0.865	0.430	0.468
	PF	0.693	0.235	0.774	0.259
	SC	1.082	0.507	0.511	-1.407

Note: Cell entries  $(i, j)$  where  $i$  indexes row and  $j$  indexes column, give the percentage change in the market share of brand  $i$  corresponding to a percentage change in the price of brand  $j$ .



## Appendix: Proof of Theorem 4.1

In this appendix we present a mathematical proof of Theorem 4.1.

**Theorem 4.1:** *Assume that the same hyperprior distribution,  $\pi(\bar{\theta}, D, \Sigma)$ , is used in the estimation of the independent samples and consumer panel demand systems. When an infinite number of consumers make purchase decisions in each period, the posterior distributions of the parameters of the model  $f(\bar{\theta}, D, \Sigma | S, X)$  under both demand systems (independent samples and panel of consumers) are equal.*

**Proof:** First recall from Proposition 4.1 that when  $N$  approaches infinity, the market shares for each period  $t$  under both demand systems (conditional on  $\bar{\theta}$ ,  $D$  and  $\xi_t$ ) converge to their expected values. Moreover, these expected values under both demand systems are also equivalent and they are given by the following expression:

$$E(s_{jt} | \bar{\theta}, D, \xi_t) = \int \frac{e^{\theta' x_{jt} + \xi_{jt}}}{\sum_{k=1}^J e^{\theta' x_{kt} + \xi_{kt}}} \phi(\theta; \bar{\theta}, D) d\theta. \quad (17)$$

Based on the argument in Berry (1994), given a fixed value of  $D$  and  $\bar{\theta}$  and an infinite number of consumers  $N$ , there exists a unique value of  $\xi_t$  such that the observed market shares ( $s_t$ ) are equal to their expected values:

$$s_{jt} = E(s_{jt} | \bar{\theta}, D, \xi_t), \quad j = 1, \dots, J. \quad (18)$$

Therefore, given  $\bar{\theta}$ ,  $D$  and the observed data  $s_t$ , the value of  $\xi_t$  that solves equation (18) is the same under both demand systems. Furthermore, based on this one-to-one correspondence between the vectors of market shares  $s_t$  and common demand shocks  $\xi_t$ , it is possible to derive the likelihood of the market share data by evaluating the density of  $\xi_t$  (i.e.,  $\phi(\xi_t; 0, \Sigma)$ ) and multiplying this density by the determinant of the corresponding Jacobian matrix. Therefore, to prove this theorem it suffices to show that the the Jacobian matrix of the transformation takes the same values under both demand systems.

First, let's denote by  $J_t$  the Jacobian matrix of the transformation of demand shocks into market shares (i.e., the  $(j, k)$  element of  $J_t$  is equal to  $\frac{ds_{jt}}{d\xi_{kt}}$ ). In addition, denote by  $J'_t$  the Jacobian matrix of the transformation of market shares into demand shocks (i.e., the  $(j, k)$  element of  $J'_t$  is equal to  $\frac{d\xi_{kt}}{ds_{jt}}$ ). It is easy to verify that  $J_t = (J'_t)^{-1}$ . Consequently, if the elements of  $J'_t$  take the same values under both demand systems, then  $J_t$  will also be equal under both systems. To verify this, recall from equation (17) that the market share of alternative  $j$  in period  $t$  under both systems when  $N$  approaches infinity is equal



to:

$$E(s_{jt}|\bar{\theta}, D, \xi_t) = \int \frac{e^{\theta' x_{jt} + \xi_{jt}}}{\sum_{k=1}^J e^{\theta' x_{kt} + \xi_{kt}}} \phi(\theta; \bar{\theta}, D) d\theta. \quad (19)$$

Therefore, the diagonal elements of  $J'_t$  under both demand systems are equal to:

$$\frac{dE(s_{jt}|\bar{\theta}, D, \xi_t)}{d\xi_{jt}} = \frac{d}{d\xi_{jt}} \int \frac{e^{\theta' x_{jt} + \xi_{jt}}}{\sum_{l=1}^J e^{\theta' x_{lt} + \xi_{lt}}} \phi(\theta; \bar{\theta}, D) d\theta, \quad (20)$$

$$= \int \frac{e^{\theta' x_{jt} + \xi_{jt}} \sum_{l=1}^J e^{\theta' x_{lt} + \xi_{lt}} - (e^{\theta' x_{jt} + \xi_{jt}})^2}{\left( \sum_{l=1}^J e^{\theta' x_{lt} + \xi_{lt}} \right)^2} \phi(\theta; \bar{\theta}, D) d\theta, \quad (21)$$

$$(22)$$

while the off-diagonal elements of  $J'_t$  under both systems correspond to:

$$\frac{dE(s_{jt}|\bar{\theta}, D, \xi_t)}{d\xi_{kt}} = \frac{d}{d\xi_{kt}} \int \frac{e^{\theta' x_{jt} + \xi_{jt}}}{\sum_{l=1}^J e^{\theta' x_{lt} + \xi_{lt}}} \phi(\theta; \bar{\theta}, D) d\theta, \quad (23)$$

$$= \int \frac{-e^{\theta' x_{jt} + \xi_{jt}} e^{\theta' x_{kt} + \xi_{kt}}}{\left( \sum_{l=1}^J e^{\theta' x_{lt} + \xi_{lt}} \right)^2} \phi(\theta; \bar{\theta}, D) d\theta. \quad (24)$$

This completes the proof as  $\xi_t$  and these derivatives which determine the Jacobian matrix are the same under both demand systems.

## References

- [1] ALLENBY GM, ROSSI PE. 2003. Bayesian Statistics and Marketing. *Marketing Science* **22**: 304-328.
- [2] BAJARI P, YE L. 2003. Deciding Between Competition and Collusion. *Review of Economics and Statistics* **85**: 971-989.
- [3] BEN AKIVA M, LERMAN S. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press: Cambridge, MA.
- [4] BERRY S. 1994. Estimating Discrete-Choice models of Product Differentiation. *RAND Journal of Economics* **25**: 242-262.
- [5] —, LEVINSOHN J, PAKES A. 1995. Automobile Prices in Market Equilibrium. *Econometrica* **63**: 841-890.
- [6] —, CARNALL M, SPILLER P. 2006. Airline Hubbing, Costs and Demand. In *Advances in Airline Economics, Vol. 1: Competition Policy and Anti-Trust*, Lee D (ed). Elsevier Press: Amsterdam, The Netherlands.
- [7] —, LINTON O, PAKES A. 2004. Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems. *Review of Economic Studies* **71**: 613-654.
- [8] BESANKO D, DUBÉ JP, GUPTA S. 2003. Competitive Price Discrimination Strategies in a Vertical Channel Using Aggregate Data. *Management Science* **49**: 1121-1138.
- [9] BERTO VILLAS-BOAS S. 2007. Vertical Relationships Between Manufacturers and Retailers: Inference With Limited Data. *The Review of Economic Studies*. **74**: 625-652.
- [10] BODAPATI A, GUPTA S. 2004. The Recoverability of Segmentation Structure from Store-level Aggregate Data. *Journal of Marketing Research* **41**: 351-364.
- [11] CHEN Y, YANG S. 2007. Estimating Disaggregate Model Using Aggregate Data via Augmentation of Individual Choice. *Journal of Marketing Research* (forthcoming).
- [12] CHIB S, GREENBERG E. 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**: 327-335.
- [13] CHINTAGUNTA P. 2001. Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data. *Marketing Science* **20**: 442-456.

- [14] CHRISTEN M, GUPTA S, PORTER J, STAELIN R, WITTINK D. 1997. Using Market-Level Data to Understand Promotion Effects in a Non-Linear Model. *Journal of Marketing Research* **34**: 322-334.
- [15] DUAN J, MELA C. 2006. The Role of Spatial Demand on Outlet Location and Pricing. Working paper, Fuqua School of Business, Duke University.
- [16] GELMAN A, RUBIN D. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**: 457-472.
- [17] —, CARLIN JB, STERN HS, RUBIN D. 1995. *Bayesian Data Analysis*. Chapman & Hall/CRC: London.
- [18] —, ROBERTS GO, GILKS WR. 1996. Efficient Metropolis jumping rules. In *Bayesian Statistics 5*, Bernardo JM *et al.* (eds). Oxford University Press: Oxford, UK.
- [19] HUBER J, TRAIN K. 2001. On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. *Marketing Letters* **12**: 257-267.
- [20] KIM B-D. 1995. Incorporating Heterogeneity with Store-Level Aggregate Data. *Marketing Letters* **6**: 159-169.
- [21] LIU J. 1994. The collapsed Gibbs sampler with applications to a gene regulation problem. *Journal of the American Statistical Association* **89**: 958-966.
- [22] MANCHANDA P, ROSSI P, CHINTAGUNTA P. 2004. Response Modeling with Nonrandom Marketing Mix Variables. *Journal of Marketing Research* **41**: 467-478.
- [23] MUSALEM A, BRADLOW E, RAJU J. 2007. Who's got the coupon? Estimating consumer preferences and coupon usage from aggregate data. Working Paper, Fuqua School of Business, Duke University.
- [24] NEVO A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**: 307-340.
- [25] NEWTON M, RAFTERY A. 1994. Approximating Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society (B)* **56**: 3-48.
- [26] ROMEO CJ. 2007. A Gibbs sampler for mixed logit analysis of differentiated product markets using aggregate data. *Computational Economics*. **29**: 3368.
- [27] ROSSI PE, MCCULLOCH RE, ALLENBY GM. 1996. The Value of Purchase History Data in Target Marketing. *Marketing Science* **15**: 321-340.

- [28] STEPHENS M. 2000. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society (B)* **62**: 795-809.
- [29] SUDHIR K. 2001. Competitive Pricing Behavior in the US Auto Market: A Structural Analysis. *Marketing Science* **20**: 42-60.
- [30] TANNER MA, WONG WH. 1987. The calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* **82**: 528-550.
- [31] YANG S, CHEN Y, ALLENBY GM. 2003. Bayesian Analysis of Simultaneous Demand and Supply. *Quantitative Marketing and Economics* **1**: 251-275.