

# **Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data**

Renna Jiang  
University of Chicago

Puneet Manchanda  
University of Michigan

Peter E. Rossi  
University of Chicago

January 2007  
Revised, August 2007

Keywords: random coefficient logit, aggregate share models, Bayesian analysis  
JEL Classifications: C11, M3

## **Abstract**

We present a Bayesian approach for analyzing aggregate level sales data in a market with differentiated products. We consider the aggregate share model proposed by Berry, Levinsohn and Pakes (1995) which introduces a common demand shock into an aggregated random coefficient logit model. A full likelihood approach is possible with a specification of the distribution of the common demand shock, although we demonstrate that our Bayes estimator works well even in the presence of a mis-specified shock distribution. We introduce a re-parameterization of the covariance matrix to improve the performance of the random walk Metropolis for covariance parameters. We illustrate the usefulness of our approach with both actual and simulated data. Sampling experiments show that our approach performs well relative to the GMM estimator.

The authors would like to thank Jean-Pierre Dube, Guenter Hitsch, Amil Petrin, and seminar participants at Chicago for comments. All authors would like thank the Kilts Center for Marketing, University of Chicago and Manchanda would like to thank the True North Communications Faculty Research Fund at the Graduate School of Business at the University of Chicago for research support. Email: [peter.rossi@chicagogsb.edu](mailto:peter.rossi@chicagogsb.edu).

## 1. Introduction

Empirical researchers often build demand models using aggregate level sales data since individual level data are not always available. Berry, Levinsohn, and Pakes (1995) – (hereafter BLP) introduced particularly appealing formulation in which a common demand shock is introduced into a random coefficient logit model to provide a coherent aggregate demand specification. This aggregate share model uses a logit specification at the individual level coupled with a normal distribution of parameters over individuals. A large and growing body of research employs the GMM technique due to Berry (1994) to estimate such models with aggregate data (c.f. Chintagunta, Dube, and Singh 2003, Davis 2007, Goldfarb, Lu, and Moorthy 2005, Nevo 2000, 2001, Sudhir 2001, Villas-Boas 2004,).

GMM estimators do not require distributional assumptions regarding the common demand shock. Our approach is to make one further distributional assumption concerning the common demand shock and derive the likelihood. Our model uses a normal distribution for the common demand shock. The resulting likelihood for aggregate share data is not in any closed form and may be quite irregular. Instead of relying on estimation procedures that require maximization, we consider Bayesian Markov Chain Monte Carlo methods which do not require a regular or even a smooth criterion function.

We apply our new Bayesian approach to both simulated and actual datasets. Our approach is relatively insensitive to simulation error in the estimates of integral terms in the density and Jacobian. This stands in marked contrast to the GMM approach. We conduct sampling experiments in which our Bayes estimator is shown to out-perform the GMM estimator. Even with a large number of simulation draws (200) used in the integral estimates, the Bayes estimator still out-performs the GMM method. The GMM method is based on a model with a tightly specified logit demand at the individual level and a normal distribution of heterogeneity but without distributional assumptions regarding the common demand shock. One might argue that the superior performance of the Bayes estimator is due

to the fact that an additional distribution assumption is used in formulating the likelihood function. A simulation example with a mis-specified model shows that the Bayes estimator works well even with a non-normal demand shock. This suggests that the reason for the superior sampling performance of the Bayes estimator is that it makes more efficient use of the data rather than using moment conditions chosen either by convention or convenience.

An additional benefit of the Bayesian approach is the ability to conduct inference for model parameters and functions of model parameters. A natural by-product of our MCMC simulation-based method is a way of constructing posterior distributions for any function of the model parameters. Indeed, it is possible to argue that price elasticities are a much more natural summary of the model parameters than the point estimates of utility weights and the covariance matrix of the random coefficient distribution. In contrast, the GMM-based approach for which the computation of asymptotic standard errors of the nonlinear functions of parameter estimates is often complex and computationally challenging. For example, under the GMM framework, some researchers have used bootstrap methods to obtain standard errors of price elasticity, price-cost margin or other various quantity of interest (Nevo 2001, Goldfarb et al. 2005). In the GMM framework, standard errors for these functions of model parameters require supplemental computations outside of the estimation algorithm. The Bayesian MCMC approach delivers the necessary computations in one unified computational framework.

There is a literature on Bayesian approaches to estimation with aggregate share data.<sup>1</sup> Chen and Yang (2003) propose a model without the common demand shock. Musalem et al (2006) develop a Bayesian approach designed for situations where there is a fixed, known and relatively small set of consumers over which aggregate demand is formed (see p. 25). Thus, they consider a situation in which there is a finite number of consumers in the

---

<sup>1</sup> Yang et al (2003) consider a Bayesian approach for small panels of disaggregate panel data. However, their approach does not extend to the case of where only aggregate data is available and there is a continuum of consumers.

population.<sup>2</sup> Finally, Romeo (2007) develops a Bayesian approach using the GMM criterion as the basis of likelihood. This likelihood (see (11), p. 40) does not include the Jacobian<sup>3</sup> terms required to derive the likelihood for the model considered here. There are two ways to interpret Romeo’s procedure – 1) as an approximation to the likelihood used here or 2) as the likelihood for some other unspecified model. The existing Bayes literature does not apply to aggregate share data derived from a continuum of consumers and a demand shock of a specified distribution. This is the goal of this paper.

The remainder of the paper is organized as follows. Section 2 presents the main model. Section 3 outlines the MCMC algorithm. Section 4 discusses the computation of elasticities. We briefly review GMM and related estimators in section 5. In section 6, we present three examples (two with simulated dataset and one with a real dataset) to evaluate the performance of our procedure and that of the GMM. Section 7 conducts sampling experiments based on findings from section 6. We consider the extension to instrumental variables in section 8 and conclude in section 9.

## 2. Model

We assume that the latent indirect utility that a consumer  $i$  derives from consuming product  $j$  at time  $t$  takes the following standard form,

$$U_{ijt} = f\left(X_{jt} \mid \theta^i\right) + \eta_{jt} + \varepsilon_{ijt} = X_{jt} \theta^i + \eta_{jt} + \varepsilon_{ijt} \quad (1)$$

where  $X_{jt}$  is a 1 by  $K$  vector that includes all the observed product attributes (e.g., brand intercepts and price  $P_{jt}$ ).  $\eta_{jt}$  is the aggregate demand shock common across households (some interpret this as a time-varying unobserved product attribute).  $\varepsilon_{ijt}$  is an idiosyncratic shock that is distributed *i.i.d.* as type I Extreme Value (0,1). There are  $J$  products and an

---

<sup>2</sup> They consider the case of 250 consumers in their implementation and indicate that their algorithm will have problems with a higher number such as 500 (footnote 9 p. 22).

<sup>3</sup> from the non-linear transformation of aggregate shocks,  $\eta$ , to shares

outside good, i.e., at any time  $t$ , household has the option of not buying any of the  $J$  products. As is standard in the literature, we characterize the distribution of household preferences via a normal distribution,  $\theta^i \sim N(\bar{\theta}, \Sigma)$ .

The predicted share is then obtained by integrating  $s_{ijt}$  over the distribution of  $\theta^i$ ,

$$\begin{aligned} s_{jt} &= \int s_{ijt} \phi(\theta^i | \bar{\theta}, \Sigma) d\theta^i \\ &= \int \frac{\exp(X_{jt}\theta^i + \eta_{jt})}{1 + \sum_{k=1}^J \exp(X_{kt}\theta^i + \eta_{kt})} \phi(\theta^i | \bar{\theta}, \Sigma) d\theta^i \end{aligned} \quad (2)$$

where  $\phi$  denotes the multivariate normal density. We can also write expected or predicted shares in terms of “mean utility” by using the identity,  $\theta^i = \bar{\theta} + v_i$   $v_i \sim N(\mathbf{0}, \Sigma)$ .

$$s_{jt} = \int \frac{\exp(\mu_{jt} + X_{jt}v_i)}{1 + \sum_{k=1}^J \exp(\mu_{kt} + X_{kt}v_i)} \phi(v_i | \mathbf{0}, \Sigma) dv$$

where  $\mu_{jt} = X_{jt}\bar{\theta} + \eta_{jt}$ . (2) shows that, at any time  $t$ , given the distribution of  $\theta^i$  and observed covariates  $X_t = (X'_{1t}, \dots, X'_{Jt})'$ , share  $s_t = (s_{1t}, \dots, s_{Jt})'$  is only a function of the aggregate demand shock  $\eta_t = (\eta_{1t}, \dots, \eta_{Jt})'$ . That is, aggregate shares inherit randomness solely from the aggregate demand shocks. We can therefore write the density of shares as a function of the density of the aggregate demand shocks. We denote the relationship between  $s_{jt}$  and  $\eta_t$  by  $h(\bullet)$  as follows,

$$s_{jt} = h(\eta_t | X_t, \bar{\theta}, \Sigma) \quad (3)$$

We further assume that the common demand shocks are independently distributed across all products with identical variances, i.e.  $\eta_{jt} \sim N(0, \tau^2)$ . The joint density of shares at time  $t$  can be obtained using the Change-of-Variable theorem as follows,

$$\begin{aligned}
\pi\left(s_{1t}, \dots, s_{Jt} \mid X_t, \bar{\theta}, \Sigma, \tau^2\right) &= \phi\left(h^{-1}\left(s_{1t}, \dots, s_{Jt} \mid X_t, \bar{\theta}, \Sigma\right) \mid \tau^2\right) J_{(\eta_t \rightarrow s_t)} \\
&= \phi\left(h^{-1}\left(s_{1t}, \dots, s_{Jt} \mid X_t, \bar{\theta}, \Sigma\right) \mid \tau^2\right) \left(J_{(s_t \rightarrow \eta_t)}\right)^{-1}
\end{aligned} \tag{4}$$

The likelihood is given by:

$$L(\bar{\theta}, \Sigma, \tau^2) = \prod_{t=1}^T \pi\left(s_t \mid X_t, \bar{\theta}, \Sigma, \tau^2\right) \tag{5}$$

To evaluate the likelihood, we need to invert the  $h$  function in (3) and evaluate the Jacobian ( $J$ ) in (4).

### Computing the inverse

To invert the shares, we use the share inversion method proposed by BLP. Rewrite the utility as:

$$U_{ijt} = \mu_{jt} + X_{jt} \nu_i + \varepsilon_{ijt} \tag{6}$$

where  $\mu_{jt} = X_{jt} \bar{\theta} + \eta_{jt}$ , and  $\nu_i \sim N(\mathbf{0}, \Sigma)$ . Expected shares can be written as an expected value over a normal distribution as in (2). Given shares and a value for  $\Sigma$ , we follow the iterative procedure proposed in BLP to obtain  $\mu_t = (\mu_{1t}, \dots, \mu_{Jt})'$ .

Since there is no analytical way to obtain the integral in (2), we numerically approximate it by averaging over a finite number of draws from  $N(\mathbf{0}, \Sigma)$ . We denote the number of draws as  $H$ . Ideally,  $H$  should be a large number to reduce simulation error. However, this is often computationally infeasible because we need to approximate the share integrals multiple times in every single evaluation of the objective function (likelihood or the GMM objective). Thus, any proposed estimation methodology for aggregate share models should work well with a small  $H$  and should not be very sensitive to the choice of  $H$ . The commonly used values in the literature are between 20 and 50.<sup>4</sup> We will investigate the sensitivity of both our approach and the GMM approach to the choice of  $H$ .

---

<sup>4</sup> We are aware of only two exceptions. Goettler and Shachar (2001) has individual-level panel data. In their paper, individual choice probability is obtained by integrating out an unobserved component in

## The Jacobian

Since  $s_t$  and  $\eta_t$  are all  $J$  by 1 vectors, the Jacobian at time  $t$  is given by

$$J_{(s_t \rightarrow \eta_t)} = \left\| \nabla_{\eta_t} s_t \right\| = \left\| \begin{bmatrix} \partial s_{1t} / \partial \eta_{1t} & \partial s_{1t} / \partial \eta_{2t} & \cdots & \partial s_{1t} / \partial \eta_{Jt} \\ \vdots & & & \\ \partial s_{Jt} / \partial \eta_{1t} & \cdots & \cdots & \partial s_{Jt} / \partial \eta_{Jt} \end{bmatrix}_{J \times J} \right\|, \quad (7)$$

where the matrix elements are obtained from (2) as

$$\partial s_{jt} / \partial \eta_{kt} = \begin{cases} \int -s_{ijt} s_{ikt} \phi(\theta^i | \bar{\theta}, \Sigma) d\theta^i & \text{if } k \neq j \\ \int s_{ijt} (1 - s_{ijt}) \phi(\theta^i | \bar{\theta}, \Sigma) d\theta^i & \text{if } k = j \end{cases} \quad (8)$$

It should be noted that conditional on shares, the Jacobian is a function of only  $\Sigma$ . To see this, write out the elements of the Jacobian:

$$\begin{aligned} & \int f(s_{ijt}) \phi(\theta^i | \bar{\theta}, \Sigma) d\theta^i \\ &= \int f \left( \frac{\exp(\mu_{jt} + X_{jt} v_i)}{1 + \sum_{k=1}^J \exp(\mu_{kt} + X_{kt} v_i)} \right) \phi(v_i | \mathbf{0}, \Sigma) dv_i \\ &= f(\mu_t, \Sigma | X_t) \end{aligned} \quad (9)$$

(9) shows that the Jacobian is a function of  $\mu_t = (\mu_{1t}, \dots, \mu_{Jt})'$  and  $\Sigma$ . But given  $\Sigma$  and shares  $s_t$ , the vector  $\mu_t$  can be uniquely determined by inverting  $s_t$ . That is, conditional on shares,  $\mu_t$  is a function of  $\Sigma$ . Therefore, conditional on shares, the Jacobian is only a function of  $\Sigma$  (not of  $\bar{\theta}$  or  $\tau^2$ ).

---

the standard logit choice probability formula. They use quasi-Monte Carlo methods with 1,024 draws. Davis (2007) uses 1,000 simulated draws to estimate consumer heterogeneity in tastes toward outside good.

### 3. Bayesian Inference Using MCMC

In this section, we combine the likelihood in (5) with priors and outline an MCMC sampler for this model.

#### Prior and Posterior

Independent priors of  $\bar{\theta}$  and  $\tau^2$  are specified as,

$$\begin{aligned}\bar{\theta} &\sim \text{MVN}(\bar{\theta}_0, V_{\bar{\theta}}) \\ \tau^2 &\sim \nu_0 s_0^2 / \chi_{\nu_0}^2\end{aligned}\tag{10}$$

We re-parameterize the covariance matrix in terms of the  $K(K+1)/2$  unique elements of the Cholesky root ( $K$  is the number of random coefficients). To enforce positive-definiteness, we re-parameterize in terms of the log of the diagonal elements of the root.

$$\begin{aligned}\Sigma &= U'U \\ U &= \begin{bmatrix} e^{r_{11}} & r_{12} & \cdots & r_{1K} \\ 0 & e^{r_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{K-1,K} \\ 0 & \cdots & 0 & e^{r_{KK}} \end{bmatrix}\end{aligned}\tag{11}$$

Denote  $r = \{r_{jk}\}_{j,k=1,\dots,K, j \leq k}$ . The priors on  $r$  are specified as

$$\begin{aligned}r_{jj} &\sim N(0, \sigma_{r\_diag}^2) \quad \text{for } j = 1, \dots, K \\ r_{jk} &\sim N(0, \sigma_{r\_off}^2) \quad \text{for } j, k = 1, \dots, K, j < k\end{aligned}\tag{12}$$

Note that we allow different priors on  $r_{jj}$  versus on  $r_{jk}$ . The reason is simple. They are measured on different scales:  $r_{jj}$  enters  $U$  after an exponential operation while  $r_{jk}$  does not.

In appendix A, we show the implications of  $\sigma_{r\_diag}^2$  and  $\sigma_{r\_off}^2$  on the implied priors for the components of  $\Sigma$  and for the correlation matrix. To summarize, the hyperparameters include

$$\{\bar{\theta}_0, V_{\bar{\theta}}, \sigma_{r\_diag}^2, \sigma_{r\_off}^2, \nu_0, s_0^2\}.$$

The joint posterior of all parameters is given by



$$\begin{aligned}
\pi(\bar{\theta}, r, \tau^2 | \{s_t, X_t\}_{t=1}^T) &\propto L(\bar{\theta}, r, \tau^2) \times \pi(\bar{\theta}, r, \tau^2) \\
&= \prod_{t=1}^T \left( J^{-1}(s_t, X_t, r) \prod_{j=1}^J \phi \left( \frac{h^{-1}(s_t | X_t, \bar{\theta}, r)}{\tau} \right) \right) \\
&\times |V_{\bar{\theta}}|^{-1/2} \exp \left\{ -\frac{1}{2} (\bar{\theta} - \bar{\theta}_0)' V_{\bar{\theta}}^{-1} (\bar{\theta} - \bar{\theta}_0) \right\} \\
&\times \prod_{j=1}^K \exp \left\{ -\frac{(r_{jj})^2}{2\sigma_{r\_diag}^2} \right\} \times \prod_{j=1}^{K-1} \prod_{k=j+1}^K \exp \left\{ -\frac{(r_{jk})^2}{2\sigma_{r\_off}^2} \right\} \\
&\times (\tau^2)^{-\left(\frac{\nu_0+1}{2}\right)} \exp \left\{ -\frac{\nu_0 s_0^2}{2\tau^2} \right\}
\end{aligned} \tag{13}$$

### MCMC Algorithm

Our MCMC algorithm is what is termed a “hybrid” algorithm. Draws from the joint posterior are accomplished using two sets of conditional distributions. The first set of conditionals can be implemented as a pure Gibbs sampler using standard natural conjugate regression theory. The second set (for  $\Sigma$  or  $r$ ) is implemented using a Metropolis step.

The basic two sets of conditionals are:

$$\begin{aligned}
&\bar{\theta}, \tau^2 | r, \{s_t, X_t\}_{t=1}^T, \bar{\theta}_0, V_{\bar{\theta}}, \nu_0, s_0^2 \\
&r | \bar{\theta}, \tau^2, \{s_t, X_t\}_{t=1}^T, \sigma_{r\_diag}^2, \sigma_{r\_off}^2
\end{aligned} \tag{14}$$

The first conditional draw for  $\bar{\theta}$  and  $\tau^2$  can be easily accomplished by computing  $\mu_{jt}$  given  $r$  and performing a univariate Bayes regression analysis. That is,

$$\mu_{jt} = X_{jt} \bar{\theta} + \eta_{jt}, \quad \eta_{jt} \sim N(0, \tau^2) \tag{15}$$

The second conditional draw for  $r$  is more complicated due to the presence of  $r$  in the Jacobian term. A Random-Walk Metropolis chain is used for the draw of  $r$ . Candidates for  $r$  are proposed according to the equation:

$$r^{new} = r^{old} + MVN(\mathbf{0}, \sigma^2 D_r) \tag{16}$$

where  $D_r$  is the candidate covariance matrix;  $\sigma^2$  is a scaling constant. We set the scaling matrix as the covariance matrix of draws from a short chain that was run for the purpose of calibrating the final chain. We then choose the scaling constant,  $\sigma^2$ , so as to maximize the information content (numerical efficiency) of the draw sequences. We note that, in order for the Metropolis step to run efficiently, different step sizes are needed for the diagonal and off-diagonal elements of  $\Sigma$  as the “ $r$ ” parameterization uses the log scale for some elements.

We note that one could define a RW Metropolis algorithm for the vector of the unique elements of  $\Sigma$  and reject proposed candidates that are not positive definite. It is well-known that such an approach can result in a MCMC method that get “stuck,” i.e. a large proportion of draws are rejected. This is particularly troublesome in high dimensional problems. Our re-parameterization enforces positive definiteness and no draws are rejected simply due to failure on the positive definiteness criterion. As far as we are aware, we are the first to propose this parameterization for direct Metropolis methods on unrestricted covariance matrices.

#### 4. Elasticities and the Demand Shock Distribution

Computation of price elasticities provides one important summary of the fitted aggregate demand model. This can either be as a summary of the substitution patterns revealed for this subset of goods or as part of the problem faced by a firm in setting prices. In either case, the researcher does not know the value of the common demand shock and needs to compute the expected aggregate demand, by integrating over the distribution of demand shock. Note that it would not be correct to condition on a specific value of the demand shock such as its mean of zero or a given realized value in the sample ( $\hat{\eta}_i$ ). Thus, a specification of the distribution of the demand shock will be required for the computation of price elasticities. This takes away some of the force of the argument for using a “distribution-free” method such as GMM. Ultimately, either positive or normative content will be derived from

predictive statements about aggregate demand which require an explicit model for the distribution of demand shocks. In typical GMM applications, a value of zero for the demand shock is used for computation of price elasticities. In a GMM approach, one could sum over the empirical distribution of realized demand shocks. The statistical properties of this estimator of the elasticity of expected demand under an unknown common shock distribution are difficult to assess.

We employ the following definition of aggregate demand price elasticity:

$$\begin{aligned}\varepsilon_{jk}(\underline{X}|\bar{\theta}, \Sigma, \tau) &\equiv \frac{\partial \ln S_j(\underline{X}|\bar{\theta}, \Sigma, \tau)}{\partial \ln P_k} \\ &= \frac{P_k}{S_j(\underline{X}|\bar{\theta}, \Sigma, \tau)} \cdot \iint_{\theta^i, \eta} \frac{\partial}{\partial P_k} \Pr(j|\underline{X}, \theta^i, \eta) \phi(\theta^i|\bar{\theta}, \Sigma) p(\eta|\tau) d\theta^i d\eta\end{aligned}\tag{17}$$

where

$$\frac{\partial}{\partial P_k} \Pr(j|\underline{X}, \theta^i, \eta) = \begin{cases} -\theta_{price}^i \cdot \Pr(j|\underline{X}, \theta^i, \eta) \cdot \Pr(k|\underline{X}, \theta^i, \eta) & j \neq k \\ \theta_{price}^i \cdot \Pr(j|\underline{X}, \theta^i, \eta) \cdot (1 - \Pr(j|\underline{X}, \theta^i, \eta)) & j = k \end{cases}$$

and  $p(\eta|\tau)$  is the distribution of common demand shock.

## 5. GMM and Related Procedures

We briefly review the GMM method and consider a related non-linear least squares method. All GMM methods are designed to exploit some sort of orthogonality conditions. If we assume that we have a matrix of variables,  $Z_t$ , that is orthogonal to  $\eta_t$ , then the moment conditions,  $E[Z_t' \eta_t]$ , can be used to define the GMM criterion function. The number of moment conditions, denoted  $M$ , does not necessarily have to exceed the number of unique parameters,  $\dim(\bar{\theta}) + \dim(r)$ , due to the highly nonlinear relationship between  $\Sigma$  and  $\eta$ . In practice, however, more moment conditions than parameters are used in an attempt to

improve the efficiency of the GMM procedure. We use a standard two-step method outlined in appendix B.

Construction of the moment conditions depends on the researcher's views on whether or not there are potentially endogenous variables in the  $X$  matrix. In our examples, we consider the classic example of price endogeneity, in which there are concerns that the price variable in  $X$  is correlated with the common demand shock. In general, there will be a set of variables in  $X$  that are assumed to be exogenous and another set that is potentially endogenous and require some sort of exogenous source of variation. In our example below, we consider the wholesale price to be a valid instrument. In our examples and sampling experiments, we will consider both the case where instruments are required and where instruments are not required.

In either the case of fully exogenous  $X$  or where there are one or more endogenous variables, we still must supply additional moment conditions other than the assumption of zero correlation between  $\eta_t$  and the set of exogenous variables and instruments. This is required for identification of the many elements of the  $\Sigma$  matrix. In our examples, we consider a fairly standard set of moment conditions motivated by the assumption that the exogenous variables in  $X$  as well as the instruments are independent of  $\eta_t$ . We consider polynomials in these variables as well as exponential and logarithm transforms of the variables. We also include interactions between these transforms of the variables and the brand intercepts. The exact set of moments is defined below in section 6.

It is also possible to define an estimator in the spirit of a non-linear least squares (NLS) estimator. To define a "least-squares" criterion function, we recognize that given  $\Sigma$ , we can define a linear estimation equation for  $\bar{\theta}$  using the mean utilities constructed by inverting the shares

$$\hat{\mu}_{jt}(\Sigma) = X_{jt} \bar{\theta} + \eta_{jt}$$

We can define an estimator of  $\bar{\theta}$  by using either least squares or two stage least squares and then form a set of “residuals” conditional on  $\Sigma$ .

$$\hat{\eta}_{jt}(\Sigma) = \mu_{jt} - X_t \hat{\theta} \quad (18)$$

This allows us to define an estimator of  $\Sigma$  by minimizing the sum of squared “residuals.”

$$\hat{\Sigma}_{NLS} = \arg \min_{\Sigma} \sum_{j=1}^J \sum_{t=1}^T \left( \hat{\eta}_{jt}(\Sigma) \right)^2 \quad (19)$$

Here we have “concentrated out”  $\bar{\theta}$  but can easily recover it by using  $\hat{\Sigma}_{NLS}$ . Even under the assumption of normally distributed  $\eta_t$ , this is not the MLE due to the presence of the Jacobian term in the likelihood. Of course, this is also a GMM estimator where the set of moments is motivated by the derivatives of  $\hat{\eta}_{jt}(\Sigma)$  with respect to  $\Sigma$ . We will consider the sampling properties of this estimator as well in our sampling experiment in section 7.

## 6. Three Examples

We consider three examples. In the first example, we use simulated data to compare the effectiveness of our approach with the GMM based approach. We then apply our approach to a real dataset and show that the difference between the two methods persists. In the third example, we use simulate data from a highly non-normal distribution (beta) of demand shocks to show that our approach is not sensitive to misspecification of the demand shock distribution.

### A Simulated Example

We choose  $J = 3$ , sample size  $T = 300$ , and  $K = 4$  which includes 3 product specific intercepts and one price coefficient. There is one outside good. Price is generated from a Uniform (0, 1) distribution to be on the same scale as the intercepts (which are 0 or 1). The design matrix  $X$  is therefore  $JT$  by  $K$ , where the first 3 columns are blocks of identity matrices, and the

last column consists of prices. We generate the common demand shocks from  $\eta_{jt} \sim N(0, \tau^2)$  in this section.

The parameter values that we use are

$$\begin{aligned} \bar{\theta} &= (-2, -3, -4, -5) \\ \Sigma &= \begin{bmatrix} 3 & 2 & 1.5 & 1 \\ & 4 & -1 & 1.5 \\ & & 4 & -0.5 \\ & & & 3 \end{bmatrix} \\ \tau &= 1 \end{aligned} \tag{20}$$

They are chosen to reflect a large outside good share (about 85 per cent) and to allow the parameters to have different signs and magnitudes. To further ensure that the parameter values and the design matrix  $X$  that we choose are realistic, we monitor the relative magnitudes of the four variance components in the utility function:  $U_{ijt} = X_{jt} \bar{\theta} + X_{jt} v_i + \eta_{jt} + \varepsilon_{ijt}$ ,  $v_i \sim N(\mathbf{0}, \Sigma)$ . Our parameter settings are designed to capture a realistic setting in which the common shock standard deviation is about one half of that of the variation in the random utility due to the terms  $X_{jt} \bar{\theta}$  or  $X_{jt} v_i$ . In addition, the correlation in the random coefficients ensures that the aggregate cross-elasticity matrix violates the IIA property.

We compare our Bayesian procedure with the GMM procedure. In the GMM procedure, we assume that price is exogenous and assume that  $\eta$  is uncorrelated to

$$\begin{aligned} &X_t \\ &x_{4,t}^2, x_{4,t}^3, x_{4,t}^4, \ln(x_{4,t}), \exp(x_{4,t}) \\ &(x_{1,t}, x_{2,t}) \times x_{4,t} \\ &(x_{1,t}, x_{2,t}) \times \ln(x_{4,t}) \\ &(x_{1,t}, x_{2,t}) \times x_{4,t}^2 \end{aligned} \tag{21}$$

for a total of 15 moment conditions.

Both procedures require simulation-based estimates of expected share integrals. To explore sensitivity of both methods to the accuracy of these integral estimates we consider the case of  $H = 50$  and  $H = 200$ . For our MCMC method, we take 40,000 posterior draws. The chain quickly burns in after 3,000 iterations and then mixes well. We use the last 37,000 iterations to compute the Bayesian posterior mean and standard deviation.<sup>5</sup>

Table 1 presents both GMM and Bayes estimates. For both  $H = 50$  and  $H = 200$ , the GMM estimates of  $\Sigma$  are quite poor - in some cases they are off by an order of magnitude. Even for the mean  $\bar{\theta}$  vector and  $\tau$ , the GMM estimates are inferior to Bayes estimates. An important point to note is that the Bayes estimates are relatively insensitive to  $H$ , while the GMM estimates are very sensitive to  $H$ . We also computed GMM estimates with  $H = 5000$ . These are no more accurate than  $H = 200$ . For example, the GMM estimates of the diagonal elements of  $\Sigma$  for  $H = 5000$  are (2.0,0.09,2.6,2.6). Of course, this is only suggestive as this involves only one sample. In section 7, we will conduct a formal sampling experiment.

Table 2 provides elasticity estimates using (17) based on both the Bayes and GMM estimates. We can see that the Bayes estimates are closer to the elasticities computed with the true parameters and that there are large differences (a factor of two or more) between the cross-elasticities computed from GMM and Bayes estimates. The GMM estimates tend to yield underestimated cross-elasticities even when  $H = 200$ .

### **Canned Tuna Data**

In this section, we apply both methods to a real dataset. We use quantity and price data for the canned tuna category from the Dominick's Finer Foods database available at the University of Chicago GSB (<http://gsbwww.uchicago.edu/kilts/research/db/dominicks/>). We aggregate the sales data to the chain level for 338 weeks. We choose the top 3 products (Star Kist 6 oz., Chicken of the Sea 6 oz., and Bumble Bee Chunk 6.12 oz.) from this category for analysis. We define the market size as the average customer count tracked by the chain - 1,934,047 (max= 2,431,995, median= 1,957,102, min= 1,094,480) - allowing us to compute an

---

<sup>5</sup> The priors are set to very diffuse values – see appendix A for details.

“outside good” share. As in many applications to scanner data, the outside share is very large, around 95 per cent.

Table 3 shows that the behavior of the GMM and Bayes estimators is much the same for the tuna data as it is for the simulated data described above. The GMM estimates often yield implausibly small values for elements of  $\Sigma$ . The Bayes estimates are also much more stable as  $H$  varies, while the GMM estimates are very sensitive to the choice of  $H$ . For  $H=200$ , the estimates of cross-elasticities based on Bayes parameter estimates are larger than the elasticities derived from GMM estimates (see table 4).

### A Simulated Example with Mis-specified Model

One general concern for likelihood based methods in this setting is that the estimates will be inconsistent if the model is mis-specified with respect to the distribution of the aggregate demand shock. In this section, we deliberately mis-specify our model and examine the obtained results.

In the data generating step, we follow the same procedure as in first simulated example except that now we generate the unobserved product attributes from  $\eta_{jt} \sim \text{Beta}[0.5, 0.5] - 0.5$ .<sup>6</sup> However, we still fit our Bayesian model under the normality assumption on  $\eta_{jt}$ . We again contrast the results with GMM, which does not make any distributional assumption on  $\eta_{jt}$ . Note that in this case, the analog for  $\tau^2$  is the variance of  $\eta_{jt}$ , 0.125.

The true parameter values are:

$$\begin{aligned}\bar{\theta} &= (2, 2, 1, -5) \\ \Sigma &= \text{diag}(1.5, 1.5, 1.5, 1.5)\end{aligned}\tag{22}$$

We also consider a diagonal matrix for  $\Sigma$  to consider the added benefits of studying the behavior of the GMM estimator in a case with no correlation or unequal variances.

---

<sup>6</sup> This is done to ensure zero expectation of  $\eta_{jt}$ .



Correlation in the random coefficient distributions is what contributes most to the flexibility of the aggregate logit model in the sense that these correlations are needed for greatest deviations from the IIA property.

Table 5 demonstrates that even mis-specified Bayes estimates are much better at recovering the parameters (especially the elements of the  $\Sigma$  matrix) than GMM. The GMM estimates of the diagonal elements of  $\Sigma$  are unrealistically small. This suggests that the GMM estimates will be very poor when there is little information in the sample to distinguish a heterogeneous logit from a homogenous model.

Figure 1 shows the histogram of the recovered values of  $\eta$  based on the GMM and Bayes estimates. While there do not seem to be big differences in the two empirical distributions, a visual inspection suggest the Bayes estimates produce a somewhat closer approximation to the true beta density function.

To summarize, across the three examples considered here, we have found consistently that the GMM approach does not perform well at recovering the heterogeneity parameter  $\Sigma$  for practical values of  $H$  (the number of simulated heterogeneity draws), and the estimate of  $\Sigma$  is sensitive to the choice of  $H$ . On the other hand, the Bayesian method works better at recovering all the model parameters at practical values of  $H$ , and it produces more stable results across different  $H$ s. Finally, there is no appreciable impact on the recovered parameters even in the case in which the distribution of the aggregate demand shock is mis-specified.

## 7. Sampling Experiment

While the simulated examples considered in section 6 are suggestive, they are not conclusive. We, therefore, carry out a sampling experiment.

### Experimental Design

The true values of  $\bar{\theta}$  and  $\Sigma$  are fixed at those used in the first example in section 6, e.g.

$$\begin{aligned}\bar{\theta} &= (-2, -3, -4, -5) \\ \Sigma &= \begin{bmatrix} 3 & 2 & 1.5 & 1 \\ & 4 & -1 & 1.5 \\ & & 4 & -0.5 \\ & & & 3 \end{bmatrix} \\ \tau &= 1\end{aligned}$$

We vary  $\tau$  and  $T$ . The values of  $\tau$  are chosen to reflect various relative magnitudes of the random components in the utility function. We choose three values of  $\tau$ : (1, 2, 0.5) while keeping the sample size,  $T$ , constant at 300, resulting in a sampling experiment with three cells: they are numbered Cell 1, Cell 2, and Cell 3, respectively. We then add a fourth cell by increasing  $T$  to 1,000 while keeping  $\tau$  at 1.

For each of the four experimental cells, we generate 50 datasets. For each of the 50 replicates for each of our four cells, we keep observables (i.e., prices) fixed and draw  $\eta$  values from the corresponding distribution  $N(0, \tau^2)$ . We then apply both the Bayesian procedure and the GMM procedure to each of our generated datasets (we use the moment conditions given by (21)).

## Results

Tables 6-8 provide MSE and Bias for each of the model parameters in each of the four experimental cells. Across a wide variety of values for  $\tau$ , the Bayes estimates dominate the GMM estimates with lower MSE. The performance of the GMM estimates is particularly poor in estimation of the diagonal elements of the  $\Sigma$  matrix. Even for the  $\bar{\theta}$  (or the “regression coefficient”) parameters, the Bayes estimates have lower bias. Figures 2-4 present the sampling distributions of the Bayes and GMM estimators in Cell 1 (labeled as the “Central Cell” in the charts). The dramatic performance differences are even more evident in these figures. One concern about the use of MSE as a performance criterion is that it can be sensitive to outliers. It is true that GMM estimates of  $\Sigma$  elements include some outliers (we took special care—see appendix B—to ensure that the GMM estimates

converged so this is not a numerical optimization problem but rather a problem with the GMM criterion).<sup>7</sup> Even if we display the sampling distribution with outliers trimmed (bottom two panels of Figure 3 and upper left panel of Figure 4), the GMM estimator is dominated by the Bayes estimator.

In Tables 6-8, we also present the sampling distribution of the NLS estimator as defined in (19). This estimator has better sampling properties than the GMM, particularly for the diagonal elements of the  $\Sigma$  matrix.

## 8. Inclusion of Instrumental Variables

Our approach can easily be extended to include instrumental variables in order to account for potential endogenous variables in the utility function. Our basic approach outlined above can easily be extended using the Gibbs sampler for the linear structural equation model outlined in Rossi et al (2005). This means that difficulties in tuning and convergence from the addition of another Metropolis step are avoided.

Denote by  $X_{jt} = \{P_{jt}, W_{jt}\}$  all the observed product attributes, where  $P_{jt}$  is a potentially endogenous characteristic (such as price) and  $W_{jt}$  includes all other observed attributes. The Bayesian analogue of an instrumental variable approach (see Yang et al. (2003) or Rossi et al. (2005), chapter 7) is a linear equation relating  $P$  to a set of instruments,  $Z_{jt}$ , and a stochastic shock  $\xi_{jt}$ , which is correlated with the demand side common shock  $\eta_{jt}$ :

$$P_{jt} = Z_{jt}\delta + \xi_{jt} \quad (23)$$

$$\begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \equiv \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12} & \Omega_{22} \end{bmatrix}\right) \quad (24)$$

We can easily derive the joint distribution of shares and prices using the standard Change-of-Variable Theorem, we obtain

---

<sup>7</sup> For Cell 1 (the central cell), the procedure GMM on average takes 0.7 hours per rep, and the Bayes procedure takes 2.2 hours per rep, both on a Pentium 4 (CPU 3 GHz, 1 GB of RAM PC) desktop.

$$\begin{aligned}
\pi(P_t, s_t | \bar{\theta}, r, \delta, \Omega) &= \pi(\xi_t, \eta_t | \bar{\theta}, r, \delta, \Omega) J_{(\xi_t, \eta_t \rightarrow P_t, s_t)} \\
&= \pi(\xi_t, \eta_t | \bar{\theta}, r, \delta, \Omega) \left( J_{(P_t, s_t \rightarrow \xi_t, \eta_t)} \right)^{-1}
\end{aligned} \tag{25}$$

and likelihood

$$L(\bar{\theta}, r, \delta, \Omega) = \prod_{t=1}^T \pi(P_t, s_t | \bar{\theta}, r, \delta, \Omega)$$

The key to obtaining the likelihood is the Jacobian,

$$J_{(P_t, s_t \rightarrow \xi_t, \eta_t)} = \begin{vmatrix} \nabla_{\xi_t} P_t & \nabla_{\eta_t} P_t \\ \nabla_{\xi_t} s_t & \nabla_{\eta_t} s_t \end{vmatrix} \tag{26}$$

where the partial derivatives are defined similarly as before. Regarding the Jacobian, first notice that  $\nabla_{\eta_t} P_t = \mathbf{0}$  since  $P_t$  does not have a direct functional relationship to the demand shock  $\eta_t$ . Second, the linear additive specification in price equation (23) implies that  $\nabla_{\xi_t} P_t = I$ . Thus, the Jacobian can be simplified into

$$J_{(P_t, s_t \rightarrow \xi_t, \eta_t)} = \begin{vmatrix} \nabla_{\xi_t} P_t & \nabla_{\eta_t} P_t \\ \nabla_{\xi_t} s_t & \nabla_{\eta_t} s_t \end{vmatrix} = \begin{vmatrix} I & \mathbf{0} \\ \nabla_{\xi_t} s_t & \nabla_{\eta_t} s_t \end{vmatrix} = \|\nabla_{\eta_t} s_t\| = J_{(s_t \rightarrow \eta_t)} \tag{27}$$

which is the same as the Jacobian obtained for a model without instruments. That is, the additional instrumental variable parameters do not affect the Jacobian. The likelihood therefore is

$$L(\bar{\theta}, r, \delta, \Omega) = \prod_{t=1}^T \left( J^{-1}(s_t, P_t, W_t, r) \prod_{j=1}^J \phi \left( \begin{bmatrix} \xi_{jt} = P_{jt} - Z_{jt} \delta \\ \eta_{jt} = h^{-1}(s_t | P_t, W_t, \bar{\theta}, r) \end{bmatrix} \middle| \Omega \right) \right) \tag{28}$$

Similar to the demand model without instrumental variables, conditional on shares, the Jacobian is only a function of  $r$ , not  $\bar{\theta}$ ,  $\delta$ , or  $\Omega$ . Further, given shares and  $r$ , we can recover  $\mu_{jt}$  and the original demand-supply system then reduces to a model of Bayes linear instrumental variables (Chapter 7 of Rossi et al. 2005). In this sense, the Bayes approach to endogeneity is similar in spirit to the linearization approach of Berry (1994). However, the

Bayes approach properly accounts for uncertainty in the estimates of  $r$  (that is, we repeatedly alternate between drawing  $r$  and inferring about the balance of the parameters given  $r$ ). Even if mean utility were directly observed, there can be severe finite sample problems with instrumental variable estimators.

$$\begin{aligned} P_{jt} &= Z_{jt}\delta + \xi_{jt}, \\ \mu_{jt} &= [W_{jt}, P_{jt}] \bar{\theta} + \eta_{jt}, \end{aligned} \quad \begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \right) \quad (29)$$

The prior specifications for  $\bar{\theta}$  and  $r$  remain the same as they are in the demand model only. We use standard priors on the additional two sets of supply parameters,

$$\begin{aligned} \delta &\sim MVN(\bar{\delta}, V_{\delta}) \\ \Omega &\sim IW(\nu_0, V_{\Omega}) \end{aligned} \quad (30)$$

Combining the likelihood with the priors, we have the following two sets of conditionals:

$$\begin{aligned} \bar{\theta}, \delta, \Omega | r, \{s_t, P_t, W_t, Z_t\}_{t=1}^T, \bar{\theta}_0, V_{\bar{\theta}}, \bar{\delta}, V_{\delta}, \nu_0, V_{\Omega} \\ r | \bar{\theta}, \delta, \Omega, \{s_t, P_t, W_t, Z_t\}_{t=1}^T, \sigma_{r\_diag}^2, \sigma_{r\_off}^2 \end{aligned} \quad (31)$$

The first conditional can be accomplished by pure Gibbs Sampler (see chapter 7 in Rossi et al 2005 for details). The draw for  $r$  conditional on other parameters involves the calculation of the Jacobian, and is done by a Metropolis step.

### A Simulated Example

We show the effectiveness of our method using a simulated example. In the data generating step, we keep the applicable parameter settings identical to those in the first simulated example in section 6. So  $J = 3$ , sample size  $T = 300$ , and  $K = 4$  which includes 3 brand specific intercepts and price. Instruments for price include  $J$  brand specific intercepts and a random variable generated from Uniform (0,1). The base parameter values that we use are:

$$\begin{aligned} \bar{\theta} &= (-2, -3, -4, -5), \quad \delta = (1, 1, 0.5, 1) \\ \Sigma &= \begin{bmatrix} 3 & 2 & 1.5 & 1 \\ & 4 & -1 & 1.5 \\ & & 4 & -0.5 \\ & & & 3 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 0.3 & 0.25 \\ & 1 \end{bmatrix} \end{aligned} \quad (32)$$

In this setup, 32.4% of the variance in prices can be explained by the instruments, i.e., the instruments are relatively strong. The correlation between the demand and price shocks is 0.46 (a moderate degree of endogeneity).

The information content of aggregate share data regarding the heterogeneity parameters is limited to begin with. Addition of instruments implies that only part of the variation in  $P$  is useful for identifying  $\bar{\theta}$  and  $\Sigma$ . In this situation, we found that chain exhibits a higher level of autocorrelation than in the previous examples and we decided to run a longer chain of 200,000 draws using a burn-in of 100,000 draws. In addition, the Bayes estimates were more sensitive to  $H$  and we found it necessary to increase  $H$  to 200. Table 9 presents the GMM and Bayes estimates of the model parameters. All model parameters are recovered with a reasonable degree of precision for the Bayes approach. The GMM estimates of elements of  $\Sigma$  are far away from the true parameters. The recovery of the mean utility parameters using GMM appear to be worse than the recovery using the Bayes approach.

### **A Sampling Experiment**

To establish that the results from the example considered above are representative, we conducted a sampling experiment. The true parameter values and the instruments are fixed at those used in the previous simulated IV example. We generate 50 datasets by first simulating 50 pairs of price and demand shocks and then calculating the implied prices and shares. We then apply both the Bayesian and GMM procedure to each dataset setting  $H$  to 200. Table 10 provides the MSE and bias for each parameter. The table shows that the GMM estimates have a large MSE and bias. In all cases, the Bayes estimates have lower MSE and, in most cases, lower bias.

## **9. Conclusions**

Aggregate share models derived from a random coefficient logit model and a common demand shock have become quite popular in the applied economics and marketing literatures. Aggregate data is more often available than panel data and firms (both retailers

and manufacturers routinely use aggregate share data). Aggregate share models are identified by variation across markets or across time. This means that there will seldom be a large sample of share observations. In addition, the empirical identification of random coefficient models with aggregate data is often tenuous. This compels interest in efficient methods of estimation. Until now, GMM methods have dominated the literature. Little is known about the sampling properties of GMM estimators except in isolated simulated examples.

We propose a full likelihood-based approach for estimation of an aggregate share model with and without instruments. Our Bayesian approach makes one more distributional assumption about the common demand shock. While this may make some researchers uncomfortable, it is important to note that the same researchers are comfortable making parametric assumptions about the other random components of the model – they usually assume a logit model with a normal distribution of heterogeneity. These assumptions are often driven by convenience. Moreover, we argue that those investigators interested in the elasticity of expected aggregate demand may require a distributional assumption regarding the demand shock.

Sampling experiments show that our Bayesian method is not only practical but that it performs better than GMM on all dimensions. The GMM estimates of the random coefficient variances can exhibit substantial downward bias, resulting in small cross-elasticities.

## References

- Berry, Steven (1994), "Estimating Discrete-choice Models of Product Differentiation", *RAND Journal of Economics*, 25(2), 242-262
- Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 63(4), 841-890
- Berry, Steve, Oliver Linton, and Ariel Pakes (2004). "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, 71, 613-654.
- Besanko, Gupta, and Jain (1998), "Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework," *Management Science*, 44, 11, 1533-1547
- Chen, Yuxin and Sha Yang (2003), "Estimating Disaggregate Model Using Aggregate Data via Augmentation of Individual Choice," *Journal of Marketing Research*, forthcoming.
- Chevalier, Judith, Anil Kashyap and Peter Rossi (2003), "Why Don't Prices Rise During Periods of Peak Demand? Evidence from scanner data," *American Economic Review*, 93, 15-37
- Chintagunta, Pradeep (2000), "A Flexible Aggregate Logit Demand Model," working paper, University of Chicago.
- Chintagunta, Dube, and Singh (2003), "Balancing Profitability and Customer Welfare in a Supermarket Chain," *Quantitative Marketing and Economics*, 1, 111-147
- Chu, Junhong, Chintagunta, and Vilcassim (2005), "Assessing the Economic Value of Distribution Channels: An Application to the PC Industry," *Journal of Marketing Research*, forthcoming.
- Davis, Peter (2007), "Spatial Competition in Retail Markets: Movie Theaters", *RAND Journal of Economics*, forthcoming.
- Dong, Xiaojing, Puneet Manchanda, and Pradeep Chintagunta (2006). "Quantifying the Benefits of Individual Level Targeting In the Presence of Firm Strategic Behavior", working paper, Santa Clara University.
- Goettler, Ronald and Ron Shachar (2001), "Spatial Competition in the Network Television Industry", *RAND Journal of Economics*, 32, 4, 624-656
- Goldfarb, Avi, Qiang Lu, and Sridhar Moorthy (2005), "Measuring Brand Equity in an Equilibrium Framework: A Structural Approach", Working paper, 2005 QME Conference.
- Musalem, Andres, Eric Bradlow, and Jagmohan Raju (2006), "Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data," *Journal of Applied Econometrics*, forthcoming.
- Nevo, Aviv (2000), "Mergers with Differentiated Products: the Case of the Ready-To-Eat Cereal Industry," *RAND Journal of Economics*, 31, 3, 395-421



Nevo, Aviv (2001), "Measuring Market Power in the Ready-To-Eat Cereal Industry," *Econometrica*, 69, 2, 307-342

Romeo, Charles (2007), "A Gibbs Sampler for Mixed Logit Analysis of Differentiated Product Markets Using Aggregate Data," *Computational Economics*, 29: 33-68.

Rossi, Peter, Greg Allenby, and Robert McCulloch (2005). *Bayesian Statistics and Marketing*, John Wiley & Sons, Ltd.

Sudhir, K (2001), "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," *Marketing Science*, 20, 1, 42-60

Villas-Boas, M. and R. Winer (1999). "Endogeneity in Brand Choice Models", *Management Science*, 45, 1324-1338.

Villas-Boas, Sofia (2004), "Vertical Contracts Between Manufacturers and Retailers: Inference with Limited Data," UC Berkeley, CUDARE working paper

Yang, Sha, Yuxin Chen, and Greg Allenby (2003). "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1, 251-275.

**Table 1**  
GMM and Bayes Estimate, Simulated Data Example

True Parameter Values				
	$\bar{\theta}$	$\Sigma$	$\tau^2$	
	-2 -3 -4 -5	$\begin{bmatrix} 3 & 2 & 1.5 & 1 \\ & 4 & -1 & 1.5 \\ & & 4 & -0.5 \\ & & & 3 \end{bmatrix}$	1	
GMM Estimates				
$H$	Estimates			
	$\bar{\theta}$	$\Sigma$	$\tau^2$	
50	-2.70 -2.34 -3.95 -6.65	$\begin{bmatrix} 7.73 & -1.13 & 3.49 & 1.05 \\ & 2.45 & 1.83 & 1.95 \\ & & 6.32 & 0.79 \\ & & & 3.50 \end{bmatrix}$	1.21	
200	-2.09 -1.85 -4.09 -4.54	$\begin{bmatrix} 3.83 & -0.76 & 4.00 & -1.06 \\ & 0.16 & -0.70 & 0.15 \\ & & 5.67 & -2.36 \\ & & & 3.71 \end{bmatrix}$	0.88	
Bayes Estimates				
$H$	Posterior Mean (Std dev)			
	$\bar{\theta}$	$\Sigma$	$\tau^2$	
50	-1.70 (0.17) -2.53 (0.19) -3.62 (0.47) -4.42 (0.34)	$\begin{bmatrix} 2.18_{(0.66)} & 2.16_{(0.65)} & 1.11_{(0.60)} & 0.81_{(0.49)} \\ & 3.34_{(0.80)} & 0.13_{(1.03)} & 1.54_{(0.52)} \\ & & 3.52_{(1.80)} & -0.92_{(0.67)} \\ & & & 2.19_{(0.92)} \end{bmatrix}$	0.87 (0.06)	Acceptance rate: 21.0%
200	-1.84 (0.19) -2.84 (0.24) -3.72 (0.42) -4.45 (0.40)	$\begin{bmatrix} 2.32_{(0.66)} & 2.35_{(0.68)} & 1.35_{(0.80)} & 0.75_{(0.60)} \\ & 4.20_{(0.96)} & 0.31_{(1.11)} & 1.57_{(0.62)} \\ & & 3.03_{(1.21)} & -0.68_{(0.62)} \\ & & & 2.04_{(0.72)} \end{bmatrix}$	0.89 (0.06)	Acceptance rate: 35.1%

**Table 2**  
Elasticity Estimates with Simulated data

<b>True Parameters</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-1.2241	0.0909	0.0331
2	0.1424	-1.1563	0.0078
3	0.1760	0.0266	-1.9279

\*Cell  $(j, k)$  represents the percentage change in  $j$ 's market share with one percent change in  $k$ 's price.

<b>GMM, H=50</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-1.4362	0.0318	0.0620
2	0.0520	-1.5687	0.0635
3	0.2241	0.1402	-2.0245

<b>Bayes, H=50</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-1.2041	0.1268	0.0326
2	0.1646	-1.0416	0.0167
3	0.1642	0.0649	-2.0005

<b>GMM, H=200</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-1.3884	0.0260	0.0682
2	0.0403	-1.2731	0.0078
3	0.3465	0.0255	-2.2409

<b>Bayes, H=200</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-1.2450	0.1249	0.0314
2	0.1549	-1.0284	0.0145
3	0.1787	0.0666	-2.0177

**Table 3**  
GMM and Bayes Estimates, Tuna Data

GMM Estimates					
H	Estimates				
	$\bar{\theta}$	$\Sigma$			$\tau^2$
50	1.92 2.87 2.38 -13.23	$\begin{bmatrix} 0.0002 & 0.001 & -0.01 & 0.02 \\ & 0.78 & 0.89 & -2.54 \\ & & 6.33 & -4.21 \\ & & & 12.71 \end{bmatrix}$			0.48
200	-2.34 2.16 2.69 -11.66	$\begin{bmatrix} 19.46 & 1.55 & 1.68 & -4.95 \\ & 6.99 & 5.70 & -11.28 \\ & & 6.27 & -11.70 \\ & & & 22.44 \end{bmatrix}$			0.46

Bayes Estimates					
H	Posterior Mean (Std dev)				
	$\bar{\theta}$	$\Sigma$			$\tau^2$
50	0.74 (0.34) 0.58 (0.37) 0.16 (0.35) -7.91 (0.55)	$\begin{bmatrix} 2.59_{(1.19)} & 3.57_{(1.35)} & 2.78_{(1.14)} & -5.35_{(1.90)} \\ & 5.44_{(1.55)} & 4.02_{(1.41)} & -7.89_{(2.06)} \\ & & 3.58_{(1.39)} & -6.07_{(1.93)} \\ & & & 12.39_{(2.98)} \end{bmatrix}$			0.33 (0.01)
Acceptance rate: 41.7%					
200	0.72 (0.36) 0.52 (0.37) 0.12 (0.34) -7.76 (0.55)	$\begin{bmatrix} 2.38_{(1.25)} & 2.93_{(1.26)} & 2.57_{(1.24)} & -4.69_{(1.97)} \\ & 4.25_{(1.35)} & 3.45_{(1.26)} & -6.42_{(1.93)} \\ & & 3.26_{(1.28)} & -5.42_{(1.93)} \\ & & & 10.46_{(3.08)} \end{bmatrix}$			0.33 (0.02)
Acceptance rate: 40.8%					

**Table 4**  
Bayes and GMM Elasticity Estimates, Tuna data

<b>GMM, H=50</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-4.0058	0.1716	0.1308
2	0.2275	-5.4450	0.1273
3	0.0915	0.0672	-5.7804

<b>GMM, H=200</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-2.4155	0.0079	0.0074
2	0.0285	-4.4819	0.0147
3	0.0290	0.0160	-4.7686

<b>Bayes, H=50</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-2.9221	0.0189	0.0145
2	0.0377	-4.7882	0.0239
3	0.0268	0.0222	-3.4651

<b>Bayes, H=200</b>			
$\begin{smallmatrix} k \\ j \end{smallmatrix}$	1	2	3
1	-3.3712	0.0212	0.0199
2	0.0385	-4.6867	0.0230
3	0.0401	0.0256	-3.9480

**Table 5**  
GMM and Bayes estimates with Beta Distributed Demand Shocks

True Parameter Values			
$\bar{\theta}$	$\Sigma$	$\tau^2$	
2	diag[1.5 1.5 1.5 1.5]	0.125	
2			
1			
-5			

GMM Estimates			
$H$	Estimates		
	$\bar{\theta}$	$\Sigma$	$\tau^2$
50	1.77	diag[0.07 0.06 0.00 0.67]	0.095
	1.81		
	1.02		
	-3.92		
200	1.82	diag[0.06 0.15 0.02 0.69]	0.095
	1.82		
	1.03		
	-4.02		

**Bayes Estimates**  
(based on normal demand shock assumption)

$H$	Posterior Mean (Stdev)			
	$\bar{\theta}$	$\Sigma$	$\tau^2$	
50	1.86 (0.03)	diag[0.34 (0.15) 0.99 (0.21) 1.06 (0.29) 1.24 (0.21)]	0.113 (0.006)	
	1.88 (0.03)			
	0.95 (0.06)			
	-4.48 (0.08)			
200	1.89 (0.04)	diag[0.87 (0.21) 1.83 (0.35) 1.16 (0.27) 1.38 (0.22)]	0.125 (0.008)	
	1.92 (0.04)			
	0.86 (0.06)			
	-4.66 (0.11)			

**Table 6**  
MSE and Bias for Estimates of  $\tau^2$  and  $\bar{\theta}$

		MSE			Bias		
		Bayes	GMM	NLS	Bayes	GMM	NLS
$\tau^2$	Cell 1	0.03	0.09	0.09	-0.15	-0.03	-0.30
	Cell 2	0.32	4.24	1.54	-0.53	0.17	-1.23
	Cell 3	0.002	0.03	0.005	-0.04	0.03	-0.07
	Cell 4	0.03	0.15	0.09	-0.17	-0.06	-0.30
$\bar{\theta}_1$	Cell 1	0.15	0.54	0.34	0.26	0.13	0.57
	Cell 2	0.14	0.52	0.24	0.14	0.14	0.47
	Cell 3	0.16	0.48	0.31	0.32	-0.02	0.54
	Cell 4	0.16	0.37	0.31	0.26	0.15	0.54
$\bar{\theta}_2$	Cell 1	0.21	0.54	1.07	0.28	0.29	1.02
	Cell 2	0.15	3.05	1.09	0.18	0.14	1.03
	Cell 3	0.22	1.91	0.99	0.35	-0.20	0.96
	Cell 4	0.24	1.82	1.07	0.38	0.13	1.03
$\bar{\theta}_3$	Cell 1	0.35	8.51	1.67	0.30	-1.51	0.54
	Cell 2	0.42	15.36	0.59	0.23	-1.64	0.67
	Cell 3	0.53	25.85	4.71	0.46	-1.89	-0.45
	Cell 4	0.38	11.19	1.16	0.50	-0.87	0.43
$\bar{\theta}_{price}$	Cell 1	0.42	1.71	2.48	0.38	0.47	1.51
	Cell 2	0.33	3.49	3.08	0.24	0.40	1.72
	Cell 3	0.87	3.71	2.20	0.83	0.30	1.34
	Cell 4	0.32	2.11	2.50	0.35	0.72	1.50

**Table 7**  
MSE and Bias for Diagonal  $\Sigma$  elements

		MSE			Bias		
		Bayes	GMM	NLS	Bayes	GMM	NLS
$\Sigma_{11}$	Cell 1	2.19	14.89	8.06	-1.00	0.13	-2.82
	Cell 2	1.60	17.13	8.38	-0.93	-0.31	-2.88
	Cell 3	2.37	23.39	6.97	-1.26	0.69	-2.61
	Cell 4	2.04	13.73	7.88	-1.07	-0.38	-2.78
$\Sigma_{22}$	Cell 1	2.35	9.52	14.30	-0.78	-0.46	-3.75
	Cell 2	2.16	188.65	15.24	-0.54	1.52	-3.90
	Cell 3	2.62	102.33	12.97	-1.27	2.60	-3.52
	Cell 4	2.53	76.64	14.19	-1.09	0.13	-3.70
$\Sigma_{33}$	Cell 1	3.03	498.86	38.73	-0.65	10.68	-1.52
	Cell 2	4.89	1083.26	13.18	-0.32	12.35	-3.31
	Cell 3	4.78	4332.13	555.38	-1.29	20.39	5.44
	Cell 4	2.51	982.51	16.51	-1.17	8.79	-1.33
$\Sigma_{44}$	Cell 1	1.78	21.73	4.20	0.40	2.19	-0.81
	Cell 2	2.39	35.80	5.98	0.73	2.60	-2.16
	Cell 3	0.83	39.47	2.67	-0.24	2.95	-0.26
	Cell 4	2.29	22.24	3.12	0.36	1.66	-0.63



**Table 8**  
MSE and Bias for Off-Diagonal  $\Sigma$  Elements

		MSE			Bias		
		Bayes	GMM	NLS	Bayes	GMM	NLS
$\Sigma_{12}$	Cell 1	1.09	10.02	3.53	-0.67	-0.29	-1.85
	Cell 2	1.06	14.34	3.72	-0.24	-0.30	-1.92
	Cell 3	1.07	16.92	3.00	-0.64	0.04	-1.68
	Cell 4	0.87	13.84	3.55	-0.56	-1.31	-1.82
$\Sigma_{13}$	Cell 1	1.30	23.49	1.87	-0.69	0.35	-1.14
	Cell 2	1.31	70.33	1.86	-0.77	-0.96	-1.30
	Cell 3	0.80	68.98	1.31	-0.66	-1.83	-0.77
	Cell 4	0.80	19.27	1.76	-0.49	-0.46	-1.12
$\Sigma_{23}$	Cell 1	1.25	19.77	1.98	0.70	0.86	1.20
	Cell 2	1.84	218.86	1.60	0.89	3.62	1.16
	Cell 3	1.77	399.27	3.32	1.04	5.14	1.59
	Cell 4	1.01	47.68	1.88	0.56	1.75	1.21
$\Sigma_{14}$	Cell 1	0.54	5.13	1.75	-0.13	-1.09	-1.23
	Cell 2	0.56	9.62	1.47	-0.18	-1.03	-1.17
	Cell 3	0.82	7.13	1.74	-0.49	-0.88	-1.14
	Cell 4	0.36	4.18	2.81	-0.02	-1.04	-1.23
$\Sigma_{24}$	Cell 1	0.50	5.19	2.54	-0.14	-0.91	-1.49
	Cell 2	0.79	14.56	2.66	-0.37	-1.26	-1.61
	Cell 3	0.90	17.53	2.51	-0.51	-1.24	-1.37
	Cell 4	0.48	13.79	1.78	-0.01	-1.48	-1.51
$\Sigma_{34}$	Cell 1	0.73	10.05	2.00	-0.15	-2.22	-1.07
	Cell 2	0.92	28.87	0.59	-0.11	-2.82	-0.04
	Cell 3	0.84	49.86	4.40	-0.28	-2.44	-1.50
	Cell 4	0.47	31.11	1.82	0.08	-2.22	-1.09

**Table 9**  
GMM and Bayes Estimates, Simulated IV Example

**True Parameter Values**

Estimates				
$\bar{\theta}$	$\Sigma$	$\delta$	$\Omega$	
-2	$\begin{bmatrix} 3 & 2 & 1.5 & 1 \\ & 4 & -1 & 1.5 \\ & & 4 & -0.5 \\ & & & 3 \end{bmatrix}$	1	$\begin{bmatrix} 0.3 & 0.25 \\ & 1 \end{bmatrix}$	
-3		1		
-4		0.5		
-5		1		

**GMM Estimates**

$H$	Estimates				
	$\bar{\theta}$	$\Sigma$	$\delta$	$\Omega$	
200	-0.90	$\begin{bmatrix} 10^{-13} & 10^{-7} & -10^{-7} & 10^{-7} \\ & 1.78 & -0.07 & 5.90 \\ & & 0.36 & -0.65 \\ & & & 3.14 \end{bmatrix}$	1.01	$\begin{bmatrix} 0.30 & 0.22 \\ & 0.80 \end{bmatrix}$	
	-6.67		1.01		
	-2.82		0.45		
	-4.25		1.02		

**Bayes Estimates**

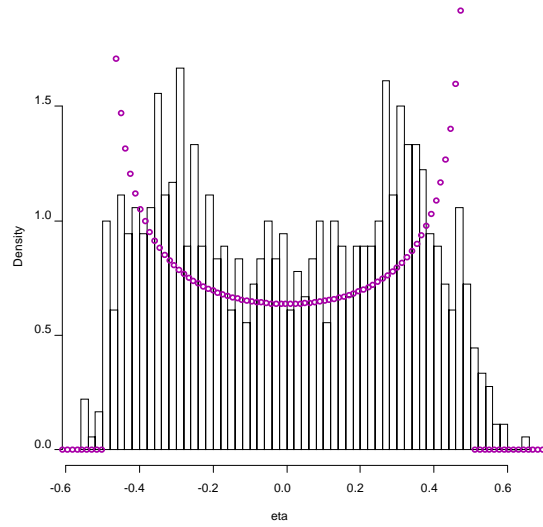
$H$	Posterior Mean (Std dev)				
	$\bar{\theta}$	$\Sigma$	$\delta$	$\Omega$	
200		$\begin{bmatrix} 1.18(0.85) & 0.98(0.63) & 0.94(0.70) & 0.54(0.18) \\ & 2.99(0.92) & -1.40(1.03) & 1.84(0.26) \\ & & 4.40(1.51) & -0.81(0.54) \\ & & & 5.41(0.84) \end{bmatrix}$	$\begin{bmatrix} 1.01(0.04) \\ 1.01(0.04) \\ 0.45(0.05) \\ 1.01(0.06) \end{bmatrix}$	$\begin{bmatrix} 0.30(0.01) & 0.25(0.04) \\ & 0.96(0.09) \end{bmatrix}$	
	-0.92 (0.41)				
	-2.17 (0.40)				
	-4.08 (0.48)				
	-5.54 (0.38)				
	Acceptance rate: 17.4%				

**Table 10**  
MSE and Bias for the IV Sampling Experiment

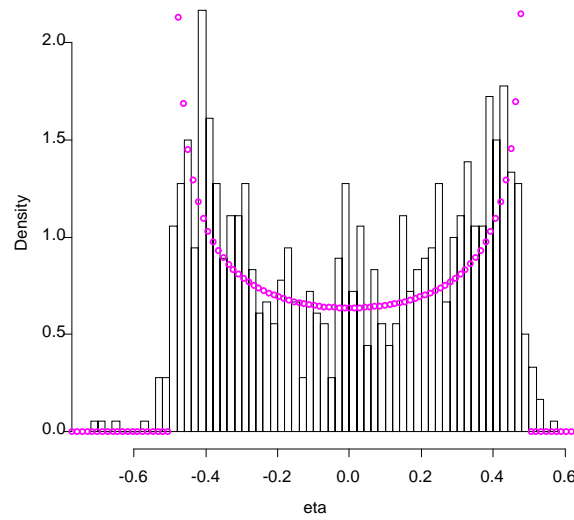
	MSE		Bias	
	Bayes	GMM	Bayes	GMM
$\bar{\theta}_1$	0.50	9.89	0.49	-0.93
$\bar{\theta}_2$	0.44	13.46	0.51	-1.28
$\bar{\theta}_3$	0.41	34.11	0.41	-2.16
$\bar{\theta}_{price}$	0.28	10	0.33	-0.02
$\Sigma_{11}$	3.82	315.49	-1.59	6.86
$\Sigma_{22}$	3.11	383.2	-1.51	8.74
$\Sigma_{33}$	3.68	6301.31	-1.30	19.09
$\Sigma_{44}$	0.75	104.68	-0.06	4.02
$\Sigma_{12}$	2.33	117.63	-1.24	1.59
$\Sigma_{13}$	1.64	82.45	-1.00	1.20
$\Sigma_{23}$	1.92	139.48	0.78	2.65
$\Sigma_{14}$	0.36	38.25	-0.25	-1.42
$\Sigma_{24}$	0.56	24.03	-0.32	-1.05
$\Sigma_{34}$	0.20	24.87	0.10	-1.89
$\delta_1$	0.002	0.002	0.003	0.001
$\delta_2$	0.002	0.002	0.002	0.001
$\delta_3$	0.002	0.002	-0.002	-0.003
$\delta_4$	0.004	0.004	-0.005	-0.002
$\Omega_{11}$	0.0002	0.0002	0.0012	-0.0003
$\Omega_{12}$	0.002	0.003	-0.02	-0.01
$\Omega_{22}$	0.03	1.28	-0.16	0.39
$Corr_{\Omega}$	0.003	0.008	-0.002	-0.062

**Figure 1**  
Recovered values of  $\eta$

**GMM**

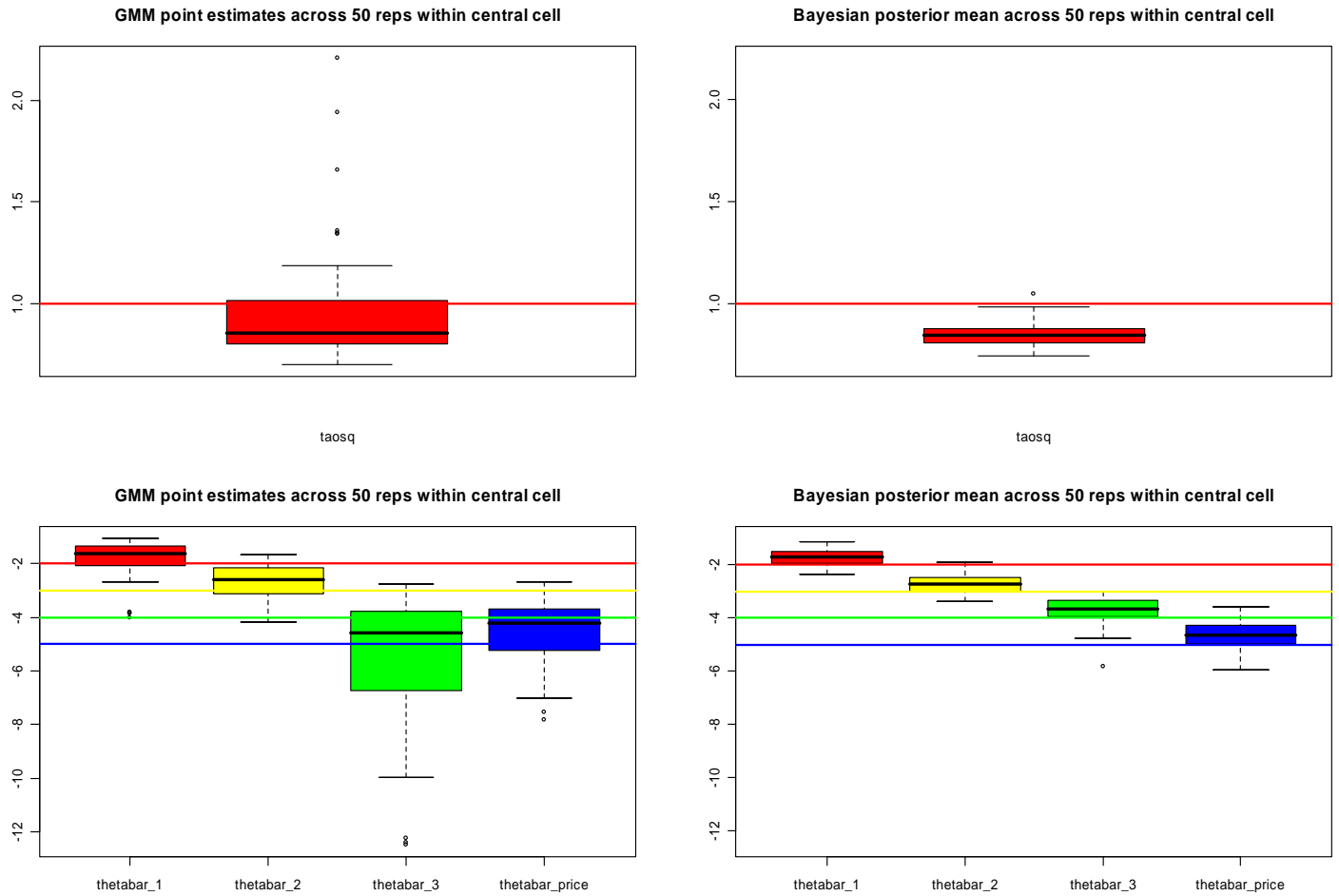


**Bayes**

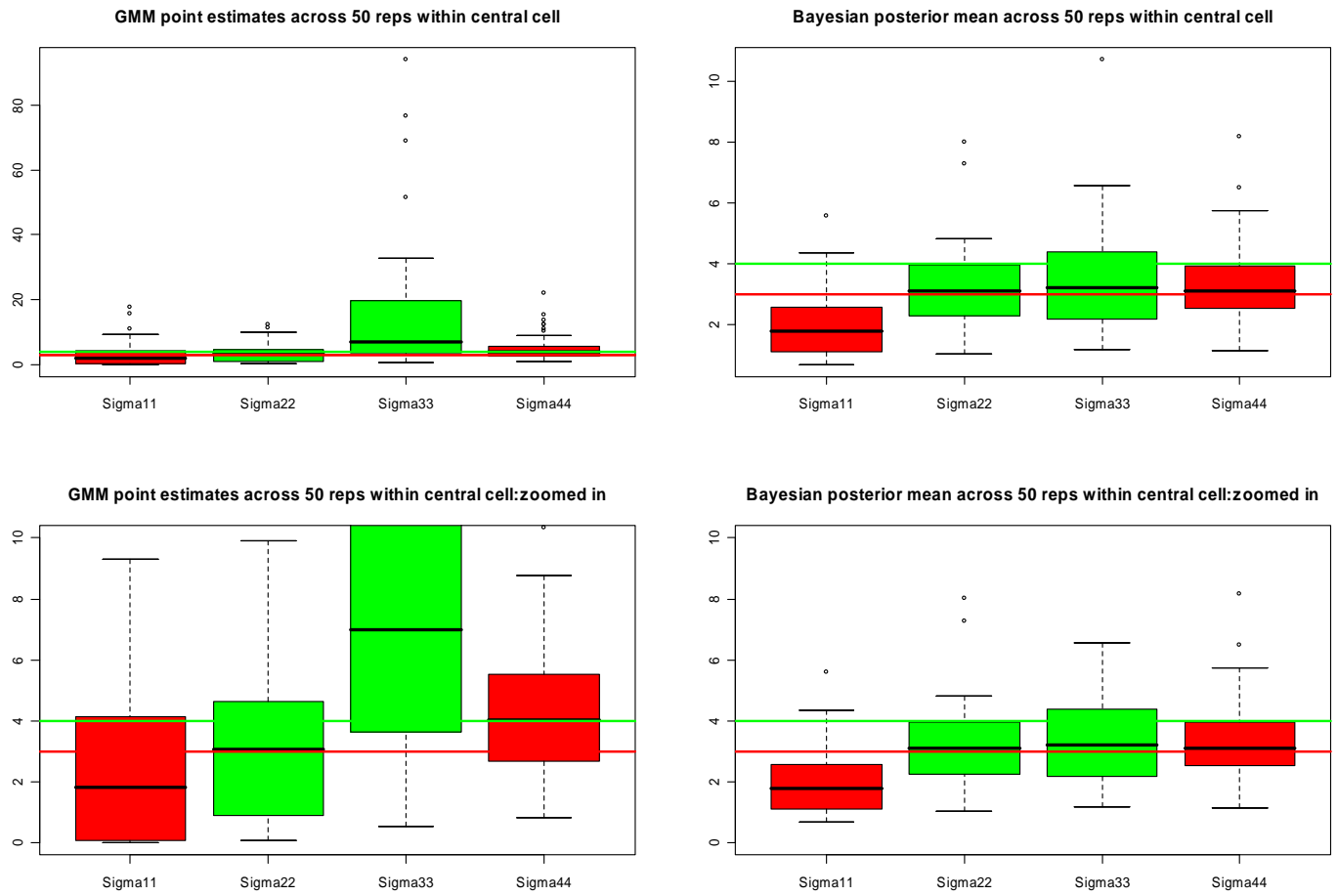


true density of  $\eta$  is shown denoted by purple circles.

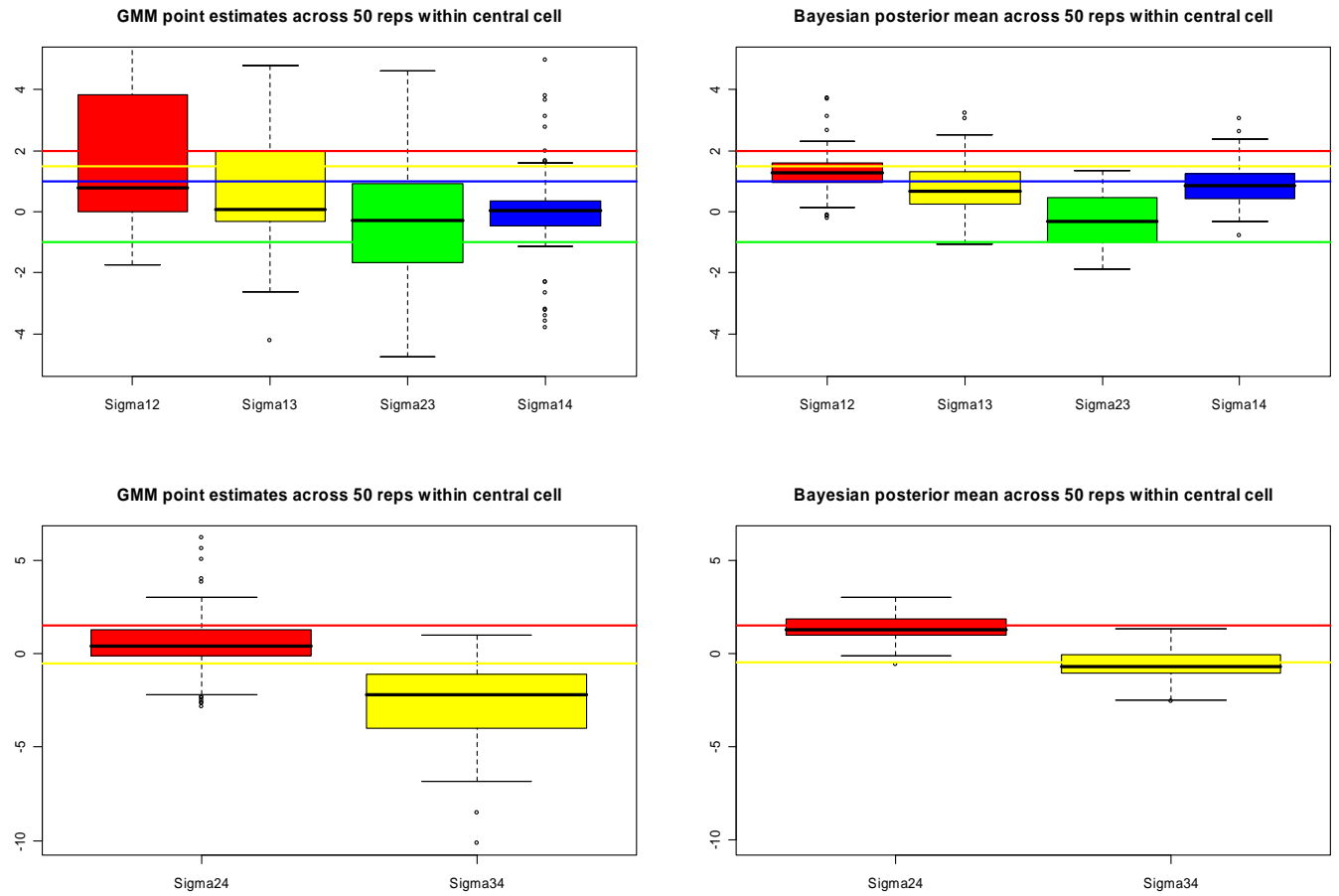
**Figure 2**  
**Sampling Distribution of GMM and Bayes Estimates of  $\tau^2$  and  $\bar{\theta}$**   
**(T=300,  $\tau=1$ )**  
 (true values and box plots are matched by colors)



**Figure 3**  
**Sampling Distribution of GMM and Bayes Estimates of Diagonal Elements of  $\Sigma$**   
**( $T=300, \tau=1$ )**



**Figure 4**  
**Sampling Distribution of GMM and Bayes Estimates of Off-Diagonal Elements of  $\Sigma$**   
**( $T=300, \tau=1$ )**



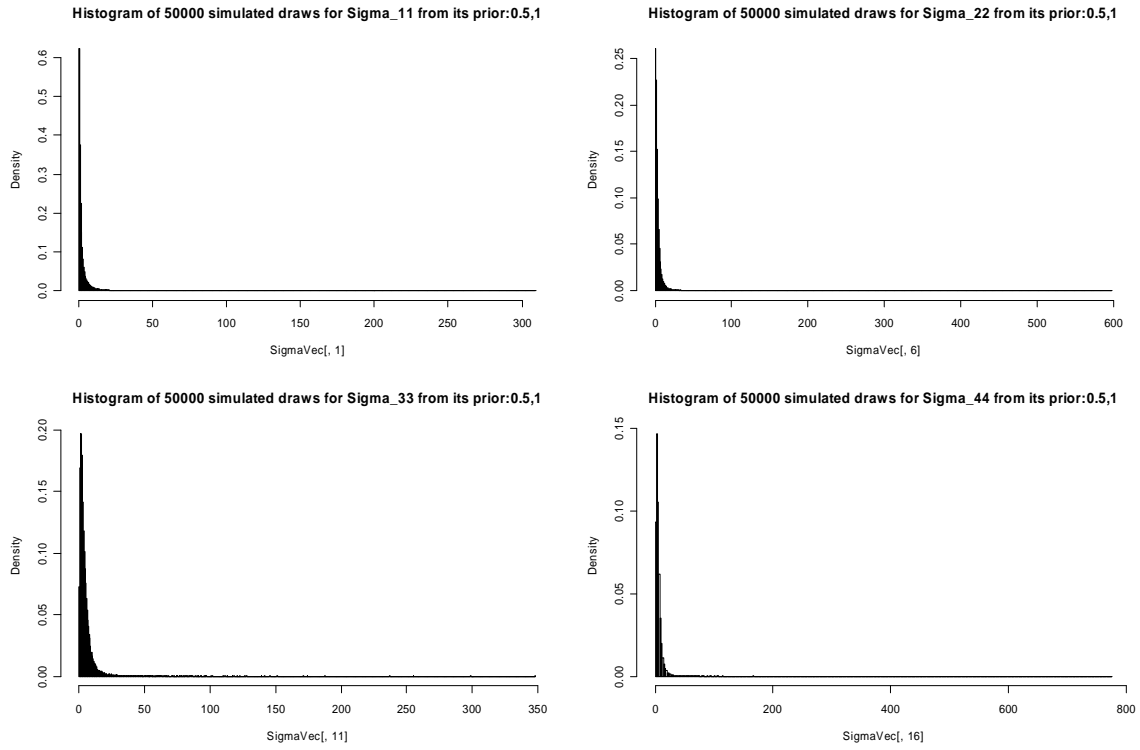
## Appendix A: Prior Settings

The hyperparameter values used in the sampling experiment and the first two examples (a simulated example and canned seafood demand data) in section 6 are as follows:

$$\begin{aligned}\overline{\theta}_0 &= \mathbf{0}, V_{\overline{\theta}} = 100I_K \\ \sigma_{r\_diag}^2 &= 0.5, \sigma_{r\_off}^2 = 1 \\ \nu_0 &= K + 1, s_0^2 = 1\end{aligned}\tag{33}$$

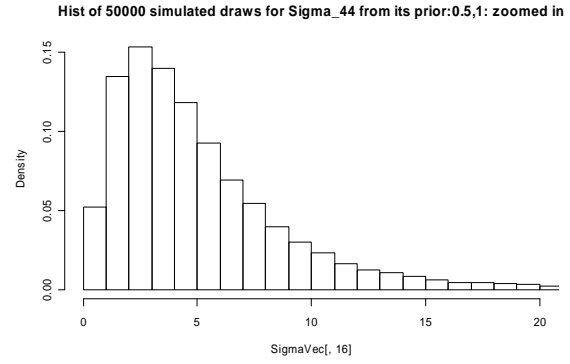
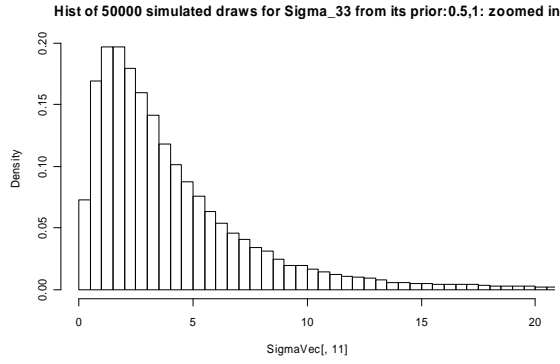
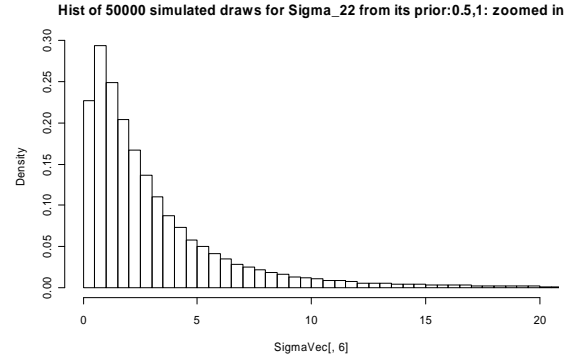
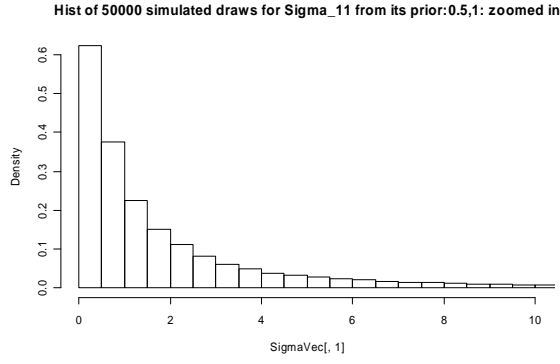
where  $I_K$  is an identity matrix of size  $K$ . Next, we plot the implied prior densities for the elements in  $\Sigma$ .

(A1). Implied prior densities for  $K$  diagonal elements in  $\Sigma$

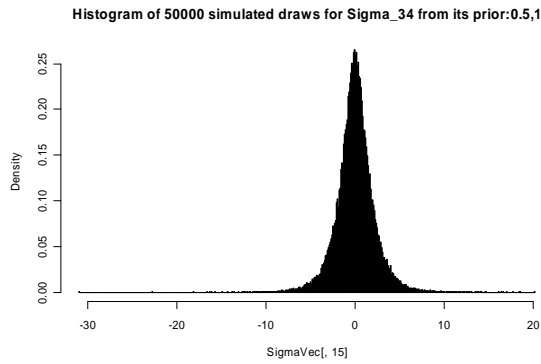
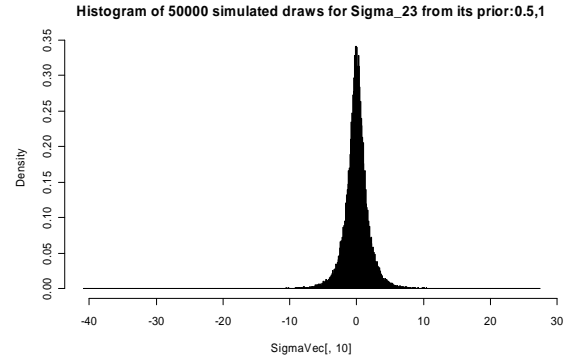
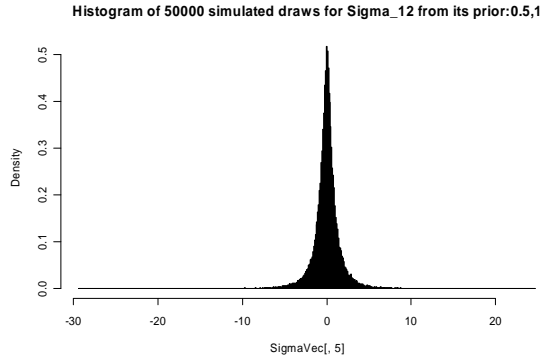


We can examine where the bulk of the density mass is:

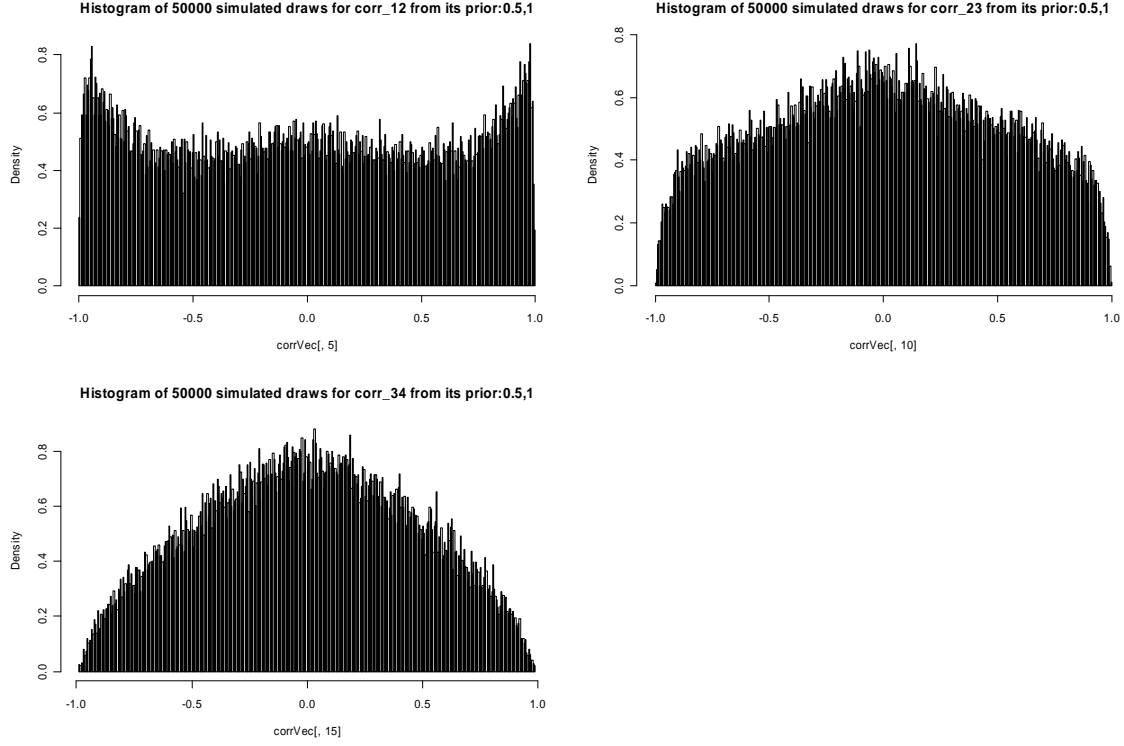




(A2). Implied prior densities for  $K - 1$  unique off-diagonal elements in  $\Sigma$



(A3). Implied prior densities for  $K - 1$  unique correlations in  $\Sigma$



In the third simulated example with mis-specified model in section 6, we use the re-parameterization scheme  $\Sigma = \text{diag}(\exp(r_1, \dots, r_K))$ , and the priors on  $r_k$  are set to be  $N(0, 10)$ , for  $k = 1, \dots, K$ .

In the simulated IV example in section 8, we keep the same priors on  $\bar{\theta}$  and  $r$  as in the demand model only (see (33)). The hyperparameters for  $\delta$  and  $\Omega$  are:

$$\bar{\delta} = \mathbf{0}, V_{\delta} = 100I_{J+1}, v_0 = K + 1, V_{\Omega} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (34)$$

## Appendix B: GMM Procedure

The theoretical moment condition is  $E[Z_t'\eta_t] = 0$ , where  $Z_t$  is  $J$  by  $M$ , and  $\eta_t$  is  $J$  by

1. Thus, there are  $M$  moments. The sample analog is  $\hat{m}_T(\bar{\theta}, \Sigma) = \frac{1}{T} \sum_{t=1}^T Z_t'(\hat{\mu}_t(\Sigma) - X_t\bar{\theta})$ .

The GMM objective is

$$\min_{\bar{\theta}, \Sigma} \hat{m}_T'(\bar{\theta}, \Sigma) A^{-1} \hat{m}_T(\bar{\theta}, \Sigma)$$

where  $A$  is the consistent estimate of the covariance matrix of the moments,  $E[(Z_t'\eta_t)(Z_t'\eta_t)']$ .  $A$  is  $M$  by  $M$ .

Take partial derivative of the GMM objective function w.r.t.  $\bar{\theta}$  and set that equal to zero, we can see that  $\bar{\theta}$  can be optimally chosen for any given  $\Sigma$  and weighting matrix  $A$ :

$$\hat{\bar{\theta}}(\Sigma) = (XZA^{-1}Z'X)^{-1} XZA^{-1}Z'\hat{\mu}(\Sigma)$$

where  $X$  is  $JT$  by  $K$ ,  $Z$  is  $JT$  by  $M$ . Thus, the GMM search is limited to  $\Sigma$ .

### Two-step GMM:

Step 1: Let  $A = \frac{1}{T} \sum_{t=1}^T Z_t'Z_t$ . Minimize the GMM objective=>obtain the residuals  $\hat{\eta}_{jt}^{(1)}$ .

Step 2: Construct a new  $A = \frac{1}{T^2} \sum_{t=1}^T Z_t'\hat{\eta}_t^{(1)}\hat{\eta}_t^{(1)'}Z_t$ , and minimize the GMM objective. After convergence, re-start the optimization routine from the converged estimates to ensure that the GMM estimates converged. Take the final converged results=> $\hat{\Sigma}_{GMM}, \hat{\bar{\theta}}_{GMM}, \hat{\eta}_{jt}^{(2)}$ .