Journal of Econometrics 149 (2009) 136–148

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bayesian analysis of random coefficient logit models using aggregate data

Renna Jiang [a], Puneet Manchanda [b], Peter E. Rossi [a,*]

[a] University of Chicago, United States

[b] University of Michigan, United States

ARTICLE INFO

Article history:
Received 20 August 2007
Received in revised form
17 November 2008
Accepted 2 December 2008
Available online 3 January 2009

JEL classification:
C11
M3

Keywords:
Random coefficient logit
Aggregate share models
Bayesian analysis

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 et al. [Berry, Steven, Levinsohn, James, Pakes, Ariel, 1995. Automobile prices in market equilibrium. Econometrica. 63 (4), 841–890], 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. We introduce a reparameterization 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 even in the presence of a mis-specified shock distribution. We view our approach as useful for those who are willing to trade off one additional distributional assumption for increased efficiency in estimation.

1. Introduction

Empirical researchers often build demand models using aggregate level sales data since individual level data are not always available. Berry et al. (1995) (hereafter BLP) introduced a 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 generalized method of moments (GMM) technique due to Berry (1994) to estimate such models with aggregate data (see, for example, Chintagunta et al. (2003), Davis (2006), Goldfarb et al. (2005), Nevo (2000, 2001), Sudhir (2001) and 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 (MCMC) methods that 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 have lower mean squared error (MSE) than the GMM estimator. 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 improved performance of the Bayes estimator is due to the fact that an additional distribution assumption is used in formulating the likelihood function. Simulations with different shock distributions and violations of the *i.i.d.* assumption show that the Bayes estimator still performs well relative to the GMM estimator. This suggests that the reason for the improved sampling performance of the Bayes estimator is that it makes more efficient use of the data.

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

* Corresponding address: University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., 60637 Chicago, IL, United States. Tel.: +1 773 702 7513; fax: +1 773 834 2081.
E-mail address: peter.rossi@chicagogsb.edu (P.E. Rossi).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.12.010

the covariance matrix of the random coefficient distribution. In contrast, the computation of asymptotic standard errors of the nonlinear functions of parameter estimates is not a natural by-product of the GMM inference procedure. 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.

We explore the adequacy of the asymptotic standard errors obtained by the GMM procedure. We find that these asymptotic standard errors understate the true sampling variance and result in confidence intervals with coverage much lower than the nominal level.

There is a literature on Bayesian approaches to estimation in aggregate share models. The existing Bayesian approaches use a data augmentation idea in which the parameters of "pseudo" consumers are added to the model. These pseudo consumer parameters are not used to integrate over the random coefficient distribution but are added to the set of parameters used in inference. Obviously, these parameters are not of interest, in and of themselves, but are merely a device to facilitate estimation. In standard data augmentation applications, the posterior, with these augmented parameters integrated out, is the same as the posterior from the likelihood function without the parameters. In the augmentation approaches for aggregate share models, this is not true. That is, the model specifies an infinite number of consumers and, thus, the existing data augmentation approaches can only approximate the correct likelihood-based inference. Moreover, this approach is limited to, at most, several hundred "pseudo" consumers in augmentation.

Chen and Yang (2007) propose a model without the common demand shock. Musalem et al. (in press) consider a model with an aggregate demand shock and an improved algorithm that uses the augmentation idea. Musalem et al. state that their method is a valid approximation method; that is, they show that, as the number of pseudo household "augmented" parameters goes to infinity, they can approximate the posterior of the model with a continuum of consumers. However, in practice, a finite and relatively small number of augmented parameters must be used. A simulation study would be required to ascertain the approximation properties of their approach.

Romeo (2007) is best viewed as a hybrid approach. He exponentiates the GMM criterion in the spirit of Chernozhukov and Hong (2005). Gallant and Hong (2007) argue that Romeo's pseudo-likelihood can be regarded as a likelihood. However, the main point is that the efficiency properties of this approach are likely to be very similar to the GMM estimator as the Romeo-style "likelihood" uses the same GMM criterion as studied here. The principal advantage of Romeo's approach is the use of a prior and the ability to simulate distributions of arbitrary functions of the model parameters (such as elasticities).

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 the GMM estimation procedure in Section 5. Section 6 conducts sampling experiments and also evaluates the adequacy of the asymptotic GMM standard errors. In Section 7, we provide an empirical example in which there is a material difference between the results obtained via the GMM and our Bayes procedure. 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:

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

where X_{jt} is a 1 by K vector that includes all the observed product attributes (e.g., brand intercepts and price P_{jt}). η_{jt} is the aggregate demand shock common across consumers/households (some interpret this as a time-varying unobserved product attribute). ε_{ijt} is an idiosyncratic shock that is distributed *i.i.d.* as a type I Extreme Value (0, 1). There are J products and an outside good, i.e., at any time t, a 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, $\theta^i \sim N\left(\bar{\theta}, \Sigma\right)$.

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

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

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

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

where $\mu_{jt} = X_{jt}\bar{\theta} + \eta_{jt}$. (2) shows that, at any time t, given the distribution of θ^i and observed covariates $X_t = \left(X_{1t}', \ldots, X_{Jt}'\right)'$, share $s_t = (s_{1t}, \ldots, s_{Jt})'$ is only a function of the aggregate demand shock $\eta_t = (\eta_{1t}, \ldots, \eta_{Jt})'$. 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 η_t by $h(\cdot)$ as follows:

$$s_{jt} = h\left(\eta_t|X_t, \bar{\theta}, \Sigma\right). \tag{3}$$

The model so far is identical to that in BLP. We add one additional assumption necessary to specify the likelihood. We assume that the common demand shocks are independently distributed across all products with identical variances, i.e. $\eta_{jt} \sim N\left(0, \tau^2\right)$. The joint density of shares at time t can be obtained using the Change-of-Variable Theorem as follows:

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

The likelihood is given by

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

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

2.1. Computing the inverse

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

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

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 Σ, we follow the iterative procedure proposed in BLP to obtain $\mu_t = \left(\mu_{1t}, \ldots, \mu_{Jt}\right)'$.

Since there is no analytical solution to 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.[1] We will investigate the sensitivity of both our approach and the GMM approach to the choice of H.

2.2. The Jacobian

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

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

where the matrix elements are obtained from (2) as

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

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

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

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

[1] We are aware of only two exceptions. Goettler and Shachar (2001) has individual level panel data. In their paper, the individual choice probability is obtained by integrating out an unobserved component in the standard logit choice probability formula. They use quasi-Monte Carlo methods with 1024 draws. Davis (2006) uses 1000 simulated draws to estimate consumer heterogeneity in tastes toward the 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.

3.1. Prior and posterior

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

$$\begin{aligned} &\bar{\theta} \sim MVN\left(\bar{\theta}_0, V_{\bar{\theta}}\right) \\ &\tau^2 \sim \nu_0 s_0^2 / \chi^2_{\nu_0}. \end{aligned} \tag{10}$$

We 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 reparameterize in terms of the log of the diagonal elements of the root.

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

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

$$\begin{aligned} &r_{jj} \sim N\left(0, \sigma^2_{r\_jj}\right) \quad \text{for } j = 1, \ldots, K \\ &r_{jk} \sim N\left(0, \sigma^2_{r\_off}\right) \quad \text{for } j, k = 1, \ldots, K, j < k. \end{aligned} \tag{12}$$

Our goal is to assess a prior which is very diffuse on both the diagonal elements of Σ as well as the correlation structure. We also would like a symmetric prior such that the prior on each diagonal element of Σ and the prior on each of the correlations are roughly the same. Equal variances of each diagonal element of Σ can be achieved if we allow different prior variances for each of the diagonal elements of the Cholesky root, r_{jj}, and one common variance for each off-diagonal element of r as in (12). In Appendix A, we explain this procedure as well as show that the implied priors for the correlations in Σ are roughly uniform. This prior is symmetric on the diagonal elements in the sense of specifying identical variances but it is not symmetric on the off-diagonal elements. However, for diffuse settings, this makes no practical difference. Care should be taken when using this prior at non-standard and non-diffuse settings. All that is required is to simulate the implied prior distribution of elements of the covariance matrix as illustrated in Appendix A. For standard diffuse settings, this step can be ignored.

This we take to be a non-dogmatic prior that allows the data to speak to the magnitude of the random effects as well as the correlation structure. It is well known that in many problems the correlation structure of Σ is very important in allowing for violations of the logit IIA assumption. By specifying a prior which is relatively diffuse, we can insure that flexible substitution patterns are possible and inferred primarily from the data. Appendix A provides the details of prior assessment for the hyperparameters $\left\{\sigma^2_{r\_jj}, j = 1, \ldots, K\right\}$ and $\sigma^2_{r\_off}$. The other hyperparameters which must be assessed are $\left\{\bar{\theta}_0, V_{\bar{\theta}}, \nu_0, s_0^2\right\}$. We employ standard diffuse settings for these hyperparameters.

The joint posterior of all parameters is given by

$$\begin{aligned} \pi\left(\bar{\theta}, r, \tau^2 \left| \{s_t, X_t\}_{t=1}^T \right.\right) &\propto L\left(\bar{\theta}, r, \tau^2\right) \times \pi\left(\bar{\theta}, r, \tau^2\right) \\ &= \prod_{t=1}^{T} \left(J^{-1}\left(s_t, X_t, r\right) \prod_{j=1}^{J} \phi\left(\frac{h^{-1}\left(s_t \left| X_t, \bar{\theta}, r \right.\right)}{\tau} \right) \right) \\ &\quad \times \left|V_{\bar{\theta}}\right|^{-1/2} \exp\left\{ -\frac{1}{2}\left(\bar{\theta} - \bar{\theta}_0\right)' V_{\bar{\theta}}^{-1} \left(\bar{\theta} - \bar{\theta}_0\right) \right\} \end{aligned}$$

$$\times \prod_{j=1}^{K} \exp\left\{-\frac{(r_{jj})^2}{2\sigma^2_{r\_jj}}\right\} \times \prod_{j=1}^{K-1}\prod_{k=j+1}^{K} \exp\left\{-\frac{(r_{jk})^2}{2\sigma^2_{r\_off}}\right\}$$
$$\times \left(\tau^2\right)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left\{-\frac{\nu_0 s_0^2}{2\tau^2}\right\}. \quad (13)$$

3.2. 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 Σ or r) is implemented using a Metropolis step.

The basic two sets of conditionals are

$$\begin{array}{l}\bar{\theta},\ \tau^2 \mid r,\ \{s_t, X_t\}_{t=1}^{T}, \overline{\theta_0}, V_{\bar{\theta}}, \nu_0, s_0^2 \\ r \mid \bar{\theta},\ \tau^2, \{s_t, X_t\}_{t=1}^{T}, \sigma^2_{r\_off}, \sigma^2_{r\_jj}.\end{array} \quad (14)$$

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

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

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

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

where D_r is the candidate covariance matrix; σ^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, σ^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 Σ as the "r" parameterization uses the log scale for some elements.

We note that one could define an RW Metropolis algorithm for the vector of the unique elements of Σ and reject proposed candidates that are not positive definite. It is well known that such an approach can result in an MCMC method that get "stuck," i.e., a large proportion of draws are rejected. This is particularly troublesome in high dimensional problems. Our reparameterization 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 description 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, therefore, 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 $\left(\hat{\eta}_t\right)$. 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 the 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 most GMM applications, a value of zero for the demand shock is used for computation of price elasticities. While there is no theoretical justification for setting the shock to zero, 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:

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

where

$$\begin{aligned}&\frac{\partial}{\partial P_k} \Pr\left(j \mid \underline{X}, \theta^i, \eta\right) \\ &= \begin{cases} -\theta^i_{\text{price}} \cdot \Pr\left(j \mid \underline{X}, \theta^i, \eta\right) \cdot \Pr\left(k \mid \underline{X}, \theta^i, \eta\right) & j \neq k \\ \theta^i_{\text{price}} \cdot \Pr\left(j \mid \underline{X}, \theta^i, \eta\right) \cdot \left(1 - \Pr\left(j \mid \underline{X}, \theta^i, \eta\right)\right) & j = k \end{cases}\end{aligned}$$

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

5. GMM procedure

We briefly review the GMM 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 η_t, then the moment conditions, $\mathrm{E}\left[Z_t'\eta_t\right]$, 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, $\dim\left(\bar{\theta}\right)+\dim\left(r\right)$, due to the highly nonlinear relationship between Σ and η. 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 requires some sort of exogenous source of variation. In our sampling experiments, we will consider both the case where instruments are required (Section 8) and where instruments are not required (Section 6).

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 η_t and the set of exogenous variables and instruments. This is required for identification of the many elements of the Σ matrix. 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 η_t. We consider polynomials in these variables as well as exponential and logarithm transforms of the variables and interactions. The exact set of moments is given in Section 6.

6. Sampling experiment

We conduct a sampling experiment to compare the sampling properties of the Bayes estimator with the GMM estimator. Since the Bayes estimator benefits from a full specification of the likelihood, we expect the Bayes estimator to perform better than the GMM estimator if the data were generated from the process underlying the likelihood. Even in the case of a correctly specified likelihood, the inefficiency of the GMM estimator relative to the Bayes estimator may be of interest. However, the potential user is also interested in the performance of the Bayes estimator relative to the GMM estimator in the case of a mis-specified model. As the aggregate share model of BLP is fully specified up to the process generating the aggregate shocks, mis-specified models consist of situations in which the data generating mechanism is not the *i.i.d.* normal model used in the formulation of the Bayes estimator. The standard justification for the GMM estimator is the belief that it will perform well under mis-specification without too much efficiency loss under a correctly specified model. However, it is important to note that, until this work, there was no real practical alternative to the GMM.

We consider five cases in our sampling experiment. These cases are designed to consider the central or base *i.i.d.* normal model and common departures from this assumption.

1. *i.i.d.* normal aggregate shock
2. Conditionally heteroskedastic shock
3. Auto-correlated shock
4. Non-normal but symmetric shock distribution
5. Non-normal and asymmetric shock distribution.

For all of the five cells in our sampling experiment, we must specify the other parameters of the model which are constant across cells. We also must specify the sample size. Our choices of these parameters are inspired by what we think might be a typical empirical application. We have in mind market share data on a set of products and an outside alternative that are collected over a time series of weeks or a collection of markets. Our X matrix includes product specific intercepts as well as a continuous variable such as log-price. We consider the case of $J = 4$ goods (3 inside and 1 outside). Thus, our X matrix consists of 3 brand intercepts and a log-price variable which is simulated as *i.i.d.* uniform (0,1). The true values of $\overline{\theta}$ and Σ also loosely correspond to those estimated in some empirical situations. The negative brand intercepts correspond to the situation in which the outside good share is large relative to any given product. The log-price coefficient is set to -5, in the range of some empirical estimates

$$\overline{\theta} = (-2, -3, -4, -5)\,. \tag{18}$$

The Σ matrix is chosen to provide a pattern of substitution quite different from the homogenous logit. We have large variance of the random coefficients and some relatively large negative correlations between the log-price coefficient and the brand intercepts.

$$\Sigma = \begin{bmatrix} 3 & 2 & 1.5 & 1 \\ & 4 & -1 & 1.5 \\ & & 4 & -0.5 \\ & & & 3 \end{bmatrix}. \tag{19}$$

We set a relatively high value of T (the number of observations) $= 300$. This is larger than many time series datasets on shares. By choosing these values, we hope to approximate a favorable case for the GMM estimator in the sense that the data are informative regarding Σ.

For each of the five experimental cells, we generate 50 datasets. For each of the 50 replicates for each of our five cells, we keep observables (i.e., prices) fixed and draw η values from the corresponding processes.

For the correctly specified model, we generate η values as *i.i.d.* $N\left(0, \tau^2\right)$ with $\tau^2 = 1$. Given the size of the slope coefficients and the variance of the X variables, this means that the observable portion of mean utility has a variance approximately 2.7 times the variance of the aggregate shock. This approximates some empirical settings and gives a relatively high information content to the data.

For the conditional heteroskedastic case, we specify

$$\begin{aligned} &\eta_{jt} \sim N\left(0, V_{jt}\right) \\ &V_{jt} = \exp\left(-.5413 + P_{jt}\right) \end{aligned} \tag{20}$$

P_{jt} is the fourth column of X and is meant to approximate a "log-price" variable. The settings of intercept $(-.5413)$ is designed to center the heteroskedastic case on the base case by insuring that $E\left[V_{jt}\right] = 1$.

The assumption of independent aggregate shocks might be called into question for time series applications where shares are measured at short time intervals such as a week or a month. To examine the violation of the independence assumption, we considered data generated from the model in which the shock corresponding to each brand follows an independent AR(1) process with the same parameters. The parameters are set as follows:

$$\begin{aligned} &\eta_{jt} = \rho\eta_{j,t-1} + u_{j,t} \\ &u_{j,t} \sim N\left(0, .4359^2\right). \end{aligned} \tag{21}$$

The variance of the innovation in the AR(1) processes is set to insure that the marginal variance of η_{jt} is still 1 as in the central cell.

To consider departures from normality, we consider drawing the shocks from a Beta distribution. We scale and translate the standard Beta distribution so that it has zero mean and appropriate variance. There are two ways in which the Beta distribution can depart from the normal. For some settings of the Beta shape parameters, the distribution is highly skewed, and for others the distribution is symmetric but "U-shaped."

For the asymmetric Beta case, we draw the shock as follows:

$$\eta_{jt} \sim 3.31305 \times Beta\,(2, 5) - .8944. \tag{22}$$

For the symmetric Beta case, we use the following draws

$$\eta_{jt} \sim 1.4142 \times Beta\,(.5, .5) - .7071. \tag{23}$$

6.1. GMM procedure and MCMC simulation settings

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

$$\begin{aligned} &X_t \\ &x_{4,t}^2, x_{4,t}^3, x_{4,t}^4, \ln\left(x_{4,t}\right), \exp\left(x_{4,t}\right) \\ &\left(x_{1,t}, x_{2,t}\right) \times x_{4,t} \\ &\left(x_{1,t}, x_{2,t}\right) \times \ln\left(x_{4,t}\right) \\ &\left(x_{1,t}, x_{2,t}\right) \times x_{4,t}^2 \end{aligned} \tag{24}$$

for a total of $M = 15$ moment conditions. In Section 8, we conduct sampling experiments for the case in which X contains some endogenous variables.

Both procedures require simulation-based estimates of expected share integrals. To explore the sensitivity of both methods to the accuracy of these integral estimates we consider the cases $H = 50$ and $H = 200$.

In our MCMC method, we take 20,000 posterior draws. The chain quickly burns in after 3000 iterations and then mixes well. We use the last 17,000 iterations to compute the Bayesian

Table 1
MSE and bias for estimates of τ^2 and $\bar{\theta}$.

		MSE		Bias	
		Bayes	GMM	Bayes	GMM
τ^2	*i.i.d. N*	0.02	0.09	−0.13	−0.03
	Hetero	0.009	0.134	−0.021	−0.011
	AR(1)	0.049	0.227	−0.172	−0.058
	Asym Beta	0.002	0.007	−0.044	−0.002
	Sym Beta	0.001	0.006	−0.03	−0.01
$\bar{\theta}_1$	*i.i.d. N*	0.11	0.54	0.22	0.13
	Hetero	0.53	0.43	−0.37	0.18
	AR(1)	0.22	0.55	0.24	0.16
	Asym Beta	0.12	0.5	0.23	0.02
	Sym Beta	0.17	0.29	0.31	0.33
$\bar{\theta}_2$	*i.i.d. N*	0.26	0.54	0.25	0.29
	Hetero	0.87	1.04	−0.52	0.25
	AR(1)	0.39	1.7	0.22	0.33
	Asym Beta	0.29	2.04	0.45	−0.14
	Sym Beta	0.25	1.52	0.33	0.10
$\bar{\theta}_3$	*i.i.d. N*	0.25	8.51	0.27	−1.51
	Hetero	2.00	12.08	−0.93	−2.02
	AR(1)	0.84	10.92	0.14	−1.46
	Asym Beta	0.41	9.39	0.50	−1.11
	Sym Beta	0.38	5.01	0.32	−1.04
$\bar{\theta}_{\text{price}}$	*i.i.d. N*	0.41	1.71	0.28	0.47
	Hetero	0.85	2.16	0.62	0.67
	AR(1)	0.59	2.39	−0.10	0.33
	Asym Beta	0.51	2.27	0.6	0.37
	Sym Beta	0.34	2.48	0.23	0.29

posterior mean and standard deviation.[2] The computing time for the Bayesian computations is roughly three times that for the GMM.

The GMM criterion for the aggregate share model of BLP has been found to be somewhat irregular, with what appear to be large "flats". The convergence criterion for termination of optimization routines can be important as well as the routine itself. We use a very stringent convergence criterion (the percentage change in the criterion function must be less than one half of machine epsilon or about 1.0^{-6}). We use both simplex and conjugate gradient methods to ensure that we have achieved an optimum. For all of our datasets, the GMM estimates are based on optimization output that successfully meets the convergence criterion.

GMM methods can also be sensitive to the criterion used to terminate the contraction mapping in the share inversion method. Often very loose criterion tolerances are used. We terminate the inversion when

$$\max\left(\left|\frac{\mu_{ij}^{k}-\mu_{ij}^{k-1}}{\mu_{ij}^{k-1}}\right|\right) < 1.0^{-6}. \tag{25}$$

In (25), k is the iteration number.

6.2. Results

Tables 1–3 provide the MSE and bias for each of the model parameters in each of the five experimental cells. In the base case of *i.i.d.* normal errors, the Bayes estimator outperforms the GMM estimator, as might be expected. It is remarkable, however, how large the MSE values are for the elements of the Σ matrix and that even for the regression parameters the Bayes estimator has an MSE one half to one fourth of that of the GMM estimator.

What is, perhaps, more remarkable is the performance of the Bayes estimator relative to the GMM estimator in the four mis-specified cells. These cells provide dramatic departures from the *i.i.d.* normal model. The performance of the Bayes estimator relative to the GMM estimator does decline somewhat for the heteroskedastic cell but does not appear to change much for any of the other conditions of autocorrelation and non-normal distributional shape, including a very skewed error distribution. It is also notable that the sampling performance of the GMM estimator degrades in the presence of heteroskedasticity and non-normality. The parameters for these cells are set to try to retain roughly the same information content in each of the cells, though information content is difficult to define in reference to a given set of moment conditions.

Table 2
MSE and bias for diagonal Σ elements.

		MSE		Bias	
		Bayes	GMM	Bayes	GMM
Σ_{11}	*i.i.d. N*	1.94	14.89	−1.04	0.13
	Hetero	11.07	25.81	1.85	0.23
	AR(1)	3.91	35.43	−0.38	0.32
	Asym Beta	2.17	66.28	−1.22	1.20
	Sym Beta	2.14	8.49	−1.19	−1.05
Σ_{22}	*i.i.d. N*	2.63	9.52	−0.70	−0.46
	Hetero	15.73	26.65	2.78	−0.33
	AR(1)	5.3	181.16	0.13	0.83
	Asym Beta	4.00	87.09	−1.71	1.96
	Sym Beta	2.34	38.38	−0.92	0.46
Σ_{33}	*i.i.d. N*	1.95	498.86	−0.62	10.68
	Hetero	40.2	566.35	3.6	12.35
	AR(1)	8.08	1927.91	0.50	15.95
	Asym Beta	3.47	601.03	−1.33	8.83
	Sym Beta	3.04	163.88	−0.42	6.04
Σ_{44}	*i.i.d. N*	2.23	21.73	0.45	2.19
	Hetero	5.12	23.11	1.16	2.33
	AR(1)	5.41	24.05	1.44	2.40
	Asym Beta	0.71	64.72	−0.16	2.73
	Sym Beta	2.42	21.58	0.50	1.91

Table 3
MSE and bias for off-diagonal Σ elements.

		MSE		Bias	
		Bayes	GMM	Bayes	GMM
Σ_{12}	*i.i.d. N*	1.24	10.02	−0.75	−0.29
	Hetero	1.75	22.0	0.46	−0.25
	AR(1)	3.14	13.49	−0.29	−0.35
	Asym Beta	0.97	23.92	−0.79	−0.84
	Sym Beta	1.22	9.84	−0.80	−0.87
Σ_{13}	*i.i.d. N*	1.15	23.49	−0.56	0.35
	Hetero	3.85	53.12	0.40	1.00
	AR(1)	2.09	49.7	−0.63	0.50
	Asym Beta	0.50	10.28	−0.54	−0.77
	Sym Beta	0.77	8.08	−0.37	−0.82
Σ_{23}	*i.i.d. N*	1.41	19.77	0.67	0.86
	Hetero	3.85	49.15	0.45	2.13
	AR(1)	3.86	173.38	0.83	4.04
	Asym Beta	1.50	15.91	0.99	1.15
	Sym Beta	1.43	17.75	0.61	1.95
Σ_{14}	*i.i.d. N*	0.54	5.13	−0.06	−1.09
	Hetero	4.05	17.47	−1.34	−1.70
	AR(1)	1.13	8.85	0.16	−0.83
	Asym Beta	0.41	12.42	−0.16	−0.53
	Sym Beta	0.78	4.95	0.19	−0.07
Σ_{24}	*i.i.d. N*	0.63	5.19	−0.08	−0.91
	Hetero	3.48	17.44	−1.02	−1.72
	AR(1)	1.42	9.47	0.08	−1.13
	Asym Beta	0.45	16.02	−0.30	−0.53
	Sym Beta	0.73	8.15	0.22	0.10
Σ_{34}	*i.i.d. N*	0.52	10.05	0.02	−2.22
	Hetero	3.01	31.25	−1.08	−3.02
	AR(1)	1.16	28.4	0.07	−1.66
	Asym Beta	0.3	27.72	−0.03	−0.44
	Sym Beta	0.74	10.43	0.09	−0.25

[2] The priors are set to very diffuse values — see Appendix A for details.

Table 4
MSE and bias for $H = 200$.

	MSE		Bias	
	Bayes	GMM	Bayes	GMM
τ^2	0.016	0.082	−0.109	−0.042
$\bar{\theta}_1$	0.10	0.42	0.13	0.1
$\bar{\theta}_2$	0.12	0.65	0.13	0.2
$\bar{\theta}_3$	0.28	3.04	0.33	−0.67
$\bar{\theta}_{\text{price}}$	0.30	1.97	0.26	0.78
Σ_{11}	1.58	12.3	−0.82	−0.12
Σ_{22}	1.58	17.12	−0.72	−0.18
Σ_{33}	2.47	66.53	−0.98	3.34
Σ_{44}	0.73	15.37	−0.01	1.37
Σ_{12}	0.72	10.76	−0.49	0.15
Σ_{13}	0.89	9.61	−0.63	−0.01
Σ_{23}	1.11	13.63	0.71	1.67
Σ_{14}	0.28	5.18	−0.13	−1.52
Σ_{24}	0.50	6.06	−0.14	−1.47
Σ_{34}	0.40	11.49	−0.05	−2.09

There is often a concern that Bayes estimators achieve better sampling performance on the basis of MSE by exploiting the bias-variance trade-off. That is, Bayes estimators can be biased. It is important to note that we explicitly set very diffuse priors so that we do not expect substantial bias. Indeed, Tables 1–3 bear this out. With the exception of the heteroskedastic case, the Bayes estimators exhibit the same or less level of bias as the GMM estimator. In the heteroskedastic case, we might expect the Bayes estimator (derived from a homoskedastic likelihood) to exhibit upward bias for the diagonal elements of Σ. It is true that the bias is greater than that of the GMM estimator for two of the four diagonal elements but this reverses for the other two. In fact, it is not the Bayes estimator that exhibits consistent bias but, rather, the GMM estimates of the off-diagonal elements of Σ. The GMM estimates are biased toward zero for the off-diagonal elements. That is, the GMM estimators tend to attenuate the correlation in the random coefficient distribution. In most cases, this would tend to bias downward cross-price elasticities of demand.

Given the relatively large MSE numbers for the GMM estimator of elements of Σ, there might be a concern that this result is driven by outliers. For this reason, we plot sampling distributions of Bayes and GMM estimators in Figs. 1–3. On the left-hand side of each figure, we show boxplots of the GMM estimates and we show Bayes estimates on the right. The true parameter values are indicated by horizontal lines of the corresponding shade of gray (or color). In Fig. 2, we magnify the plots in the bottom panels to eliminate outliers. The "box" in the boxplot is shows the interquartile range of the distribution. These plots bear out the fact that the sampling distribution of the Bayes estimators shows dramatically less variability and is located closer to the true values.

In the GMM literature, it is rare to see more than 50 or so draws used in approximating the integrals required to evaluate expected shares. It is possible that the GMM estimator would improve with a large number of draws. In Table 4, we present the results of a sampling experiment for our base cell (*i.i.d.* normal) with $H = 200$ draws. The GMM sampling properties improve somewhat with a larger number of draws. However, the Bayes sampling properties are virtually unchanged and the relative dominance of the Bayes estimator is retained. This is an important advantage of the Bayesian approach.

6.3. Adequacy of GMM asymptotic standard errors

While the sampling properties of an estimation procedure are important, it is also important to consider the quality of inference available with a given procedure. As has already been noted, this is a real advantage for modern Bayesian methods as the posterior distribution of any quantity of interest can be obtained using the posterior draws. For the GMM estimators, we must compute asymptotic standard errors and then use the delta method or parametric bootstrap to approximate the sampling distribution of functions of these parameters.

To investigate the quality of asymptotic inference available from the GMM, we computed the asymptotic standard errors of all model parameters for the central *i.i.d.* normal cell for our sampling experiment. We use the standard GMM formulas for asymptotic standard errors (see Appendix B). Following the advice generally given in the literature, we used a large number of draws in computation of the standard errors ($H = 200$).

Obviously, these standard errors depend on the estimated parameters and vary from replication to replication in the sampling experiment. To give some idea of the general properties of these standard errors, we compare the size of the median asymptotic standard error relative to the actual estimated sample standard deviation computed from the GMM estimates for each of the 50 replications. Fig. 4 provides a box plot of the ratio of the sample standard deviation for each of the parameters to the median asymptotic standard error (computed for each simulated dataset). We use the median to insulate this procedure from outliers. Fig. 4 shows that the asymptotic standard errors are about one half the size of the actual sampling standard deviation. This implies that asymptotics have not set in and are giving an overly optimistic view of the precision of estimation. This is remarkable in the case of 300 *i.i.d.* observations (this gives us a ratio of observations to parameters of about 20). The understatement of sampling error is consistent across all model parameters, not just the elements of Σ.

The consistently small standard errors should give rise to confidence intervals with dramatically lower coverage than the nominal confidence level. To assess this, we computed confidence intervals for each parameter across all 50 replications using the GMM estimate $+/-$ 1.96 asymptotic standard errors. 440 of these intervals cover the true parameter values out of the 14 (number of parameters) $\times$ 50 replications $=$ 700 possibilities (note: we do not include τ^2 as this parameter is not directly estimated by the GMM procedure). This gives a coverage rate of 63%, far below the nominal level. This is not the result of good coverage for some parameters and extremely bad coverage for others, but rather the consistent underestimation of the sampling error by the asymptotics. An interesting question, which we leave for others, is to ask what sample size would be required for asymptotics to provide a good approximation. We should note that, in order to present the most favorable possible case for asymptotics, we construct confidence intervals on r not on Σ. In our parameterization, r is unrestricted while the unique elements of Σ are constrained by positive definiteness. It is thought that asymptotics will set in faster for unrestricted parameters than for restricted parameters.

Some argue that the Bayesian procedure can also be used to construct approximate confidence intervals using the Bayes estimates and the posterior standard deviations. The correspondence between Bayes credibility regions and confidence intervals only holds asymptotically. However, we include this calculation for the sake of completeness. If we construct "95%" Bayes credibility regions, we obtain a much higher coverage rate of 81%.

7. An empirical example

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 Graduate School of Business (GSB) (http://gsbwww.uchicago.edu/kilts/research/db/dominicks/). We aggregate the sales data to the chain level for 338 weeks (we note

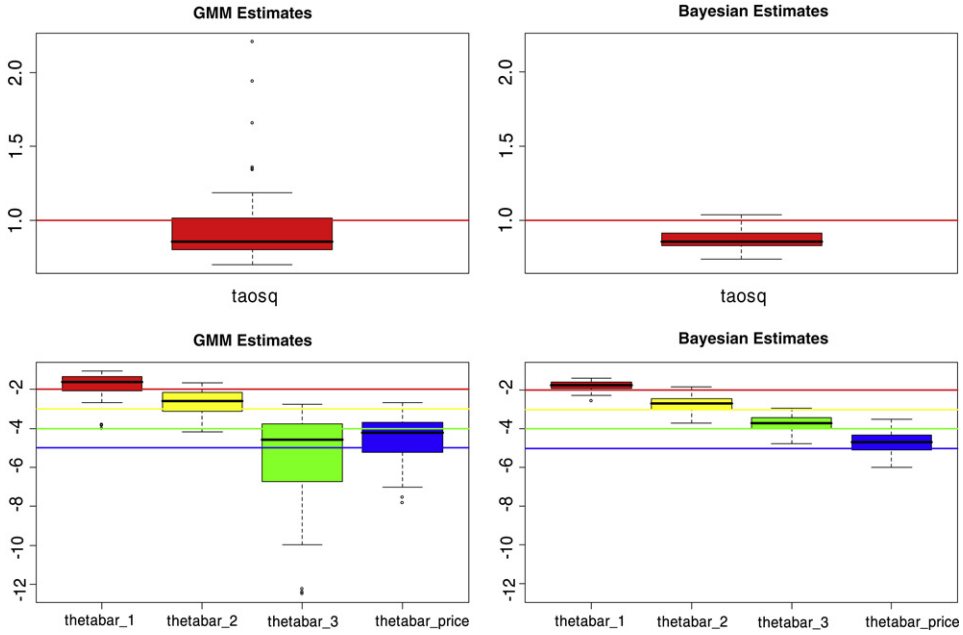


Fig. 1. Sampling distribution of GMM and Bayes estimates of τ^2 and $\bar{\theta}$.

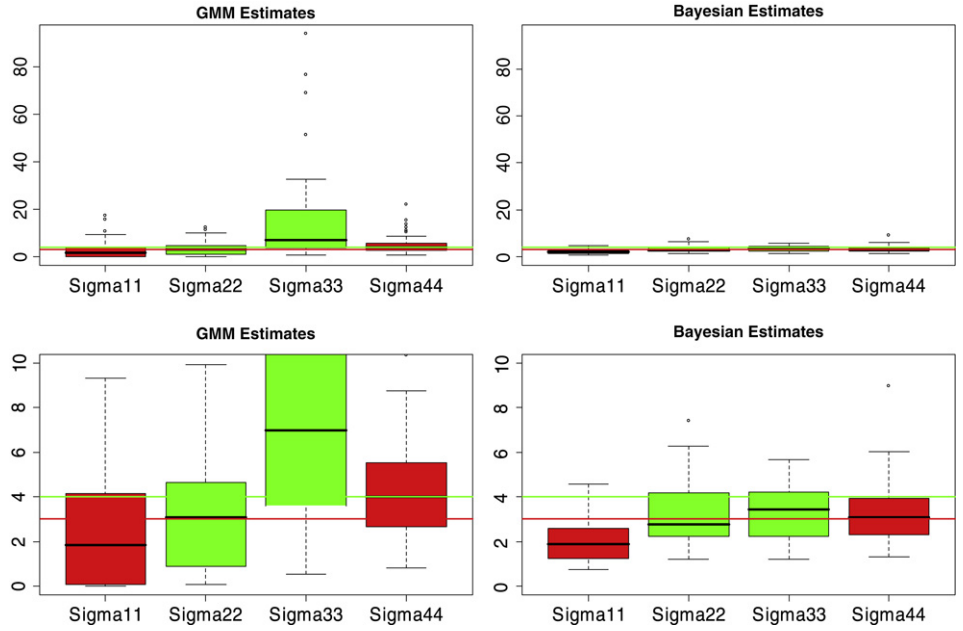


Fig. 2. Sampling distribution of GMM and Bayes estimates of diagonal elements of Σ.

that these data are unique for the length of the weekly time series). We choose the top three 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 – 1934,047 (max = 2431,995, median = 1957,102, min = 1094,480) – allowing us to compute an "outside good" share. As in many applications to scanner data, the outside share is very large, around 95%.

Table 5 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 Σ. 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 6). We hope that this empirical example shows that the results of the sampling experiments are relevant to at least some empirical applications.

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 log-price) and W_{jt} includes all other observed attributes. The Bayesian

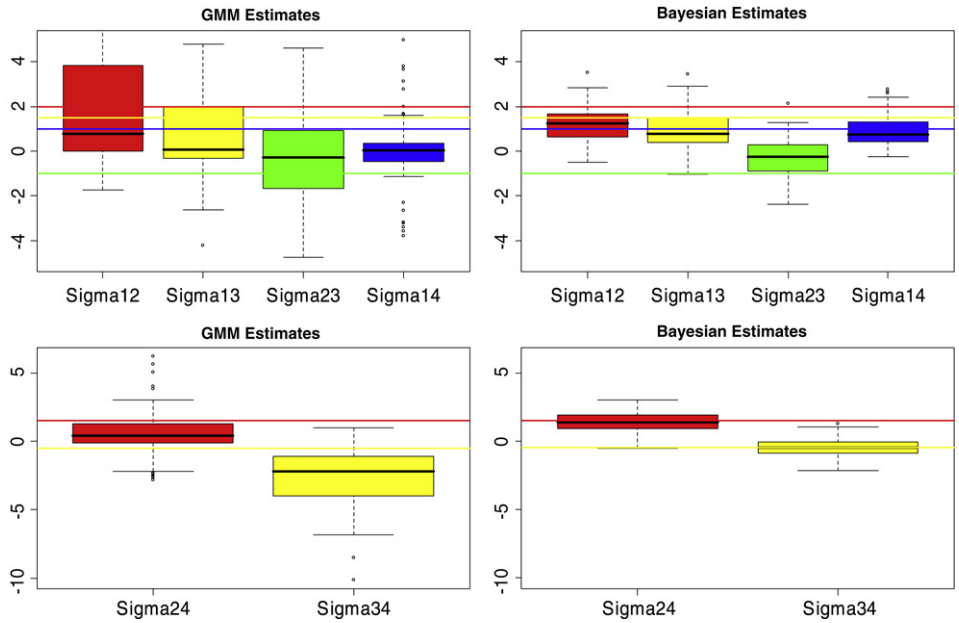


Fig. 3. Sampling distribution of GMM and Bayes estimates of off-diagonal elements.

Table 5
GMM and Bayes estimates, tuna data.

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

Table 6
Bayes and GMM elasticity (Cell (j, k) displays the percentage change the market share of product j with respect to percentage change in the price of product k) estimates, tuna data.

GMM, $H = 50$				Bayes, $H = 50$			
j	k			j	k		
	1	2	3		1	2	3
1	−4.0058	0.1716	0.1308	1	−2.9221	0.0189	0.0145
2	0.2275	−5.4450	0.1273	2	0.0377	−4.7882	0.0239
3	0.0915	0.0672	−5.7804	3	0.0268	0.0222	−3.4651
GMM, $H = 200$				Bayes, $H = 200$			
1	−2.4155	0.0079	0.0074	1	−3.3712	0.0212	0.0199
2	0.0285	−4.4819	0.0147	2	0.0385	−4.6867	0.0230
3	0.0290	0.0160	−4.7686	3	0.0401	0.0256	−3.9480

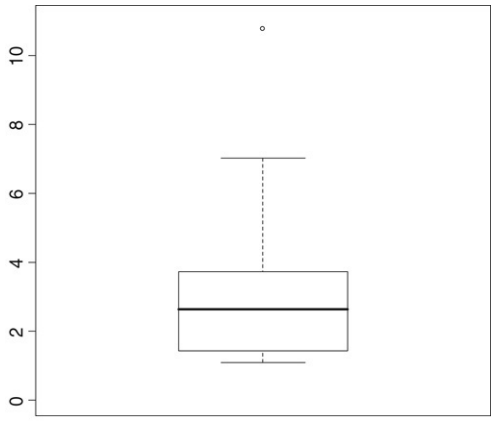


Fig. 4. Ratio of sampling standard deviation to asymptotic standard errors.

analogue of an instrumental variable approach (see Yang et al. (2003) or Rossi et al. (2005, Chapter 7)) is a linear equation relating P (here P denotes log-price) to a set of instruments, Z_{jt}, and a stochastic shock ξ_{jt}, which is correlated with the demand side common shock η_{jt}:

$$P_{jt} = Z_{jt}\delta + \xi_{jt} \tag{26}$$

$$\begin{pmatrix}\xi_{jt}\\ \eta_{jt}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Omega \equiv \begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{12} & \Omega_{22}\end{bmatrix}\right). \tag{27}$$

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

$$\begin{aligned}\pi\left(P_t, s_t \,|\overline{\theta}, r, \delta, \Omega\right) &= \pi\left(\xi_t, \eta_t \,|\overline{\theta}, r, \delta, \Omega\right) J_{(\xi_t,\eta_t\to P_t,s_t)}\\ &= \pi\left(\xi_t, \eta_t \,|\overline{\theta}, r, \delta, \Omega\right)\left(J_{(P_t,s_t\to\xi_t,\eta_t)}\right)^{-1}\end{aligned} \tag{28}$$

and the likelihood

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

The key to obtaining the likelihood is the Jacobian,

$$J_{(P_t,s_t\to\xi_t,\eta_t)} = \left\|\begin{matrix}\nabla_{\xi_t} P_t & \nabla_{\eta_t} P_t\\ \nabla_{\xi_t} s_t & \nabla_{\eta_t} s_t\end{matrix}\right\| \tag{29}$$

where the partial derivatives are defined similarly as before. Regarding the Jacobian, first notice that $\nabla_{\eta_t} P_t = 0$ since P_t does not have a direct functional relationship to the demand shock η_t. Second, the linear additive specification in log-price Eq. (26) implies that $\nabla_{\xi_t} P_t = I$. Thus, the Jacobian can be simplified to

$$\begin{aligned}J_{(P_t,s_t\to\xi_t,\eta_t)} &= \left\|\begin{matrix}\nabla_{\xi_t} P_t & \nabla_{\eta_t} P_t\\ \nabla_{\xi_t} s_t & \nabla_{\eta_t} s_t\end{matrix}\right\|\\ &= \left\|\begin{matrix}I & \mathbf{0}\\ \nabla_{\xi_t} s_t & \nabla_{\eta_t} s_t\end{matrix}\right\| = \left\|\nabla_{\eta_t} s_t\right\| = J_{(s_t\to\eta_t)}\end{aligned} \tag{30}$$

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

$$\begin{aligned}L\left(\overline{\theta}, r, \delta, \Omega\right) = \prod_{t=1}^{T}\Bigg(&J^{-1}\left(s_t, P_t, W_t, r\right)\prod_{j=1}^{J}\phi\\ &\times\left(\left[\begin{matrix}\xi_{jt} = P_{jt} - Z_{jt}\delta\\ \eta_{jt} = h^{-1}\left(s_t \,|P_t, W_t, \overline{\theta}, r\right)\end{matrix}\right]\Bigg|\,\Omega\right)\Bigg).\end{aligned} \tag{31}$$

Similar to the demand model without instrumental variables, conditional on shares, the Jacobian is only a function of r, not $\overline{\theta}$, δ, or Ω. Further, given shares and r, we can recover μ_{jt} and the original demand–supply system then reduces to a model of Bayes linear instrumental variables (see 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 remaining parameters given r). Even if the mean utility were directly observed, there can be severe finite sample problems with instrumental variable estimators.

$$\begin{aligned}&P_{jt} = Z_{jt}\delta + \xi_{jt},\\ &\mu_{jt} = \left[W_{jt}, P_{jt}\right]\bar{\theta} + \eta_{jt},\end{aligned}\quad \begin{pmatrix}\xi_{jt}\\ \eta_{jt}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Omega\right). \tag{32}$$

The prior specifications for $\overline{\theta}$ 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,

$$\begin{aligned}&\delta \sim MVN\left(\bar{\delta}, V_\delta\right)\\ &\Omega \sim IW\left(\nu_0, V_\Omega\right).\end{aligned} \tag{33}$$

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

$$\begin{aligned}&\overline{\theta}, \delta, \Omega \,|r,\ \{s_t, P_t, W_t, Z_t\}_{t=1}^{T}, \overline{\theta_0}, V_{\overline{\theta}}, \bar{\delta}, V_\delta, \nu_0, V_\Omega\\ &r\left|\overline{\theta},\ \delta, \Omega, \{s_t, P_t, W_t, Z_t\}_{t=1}^{T}, \sigma^2_{r\_jj}, \sigma^2_{r\_off}\right.\end{aligned} \tag{34}$$

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.

8.1. A sampling experiment

We show the properties of our method using a sampling experiment. In the data generating step, we keep the applicable parameter settings identical to those in the central cell of the sampling experiments reported in Section 6. Recall that $J = 3$, sample size $T = 300$, and $K = 4$ which includes three brand-specific intercepts and log-price. Instruments for the log-price include J brand-specific intercepts and a random variable generated from a Uniform (0, 1) distribution (as might be the approximate distribution of the log-wholesale price). The base parameter values that we use are

$$\overline{\theta} = (-2, -3, -4, -5), \qquad \delta = (1, 1, 0.5, 1)$$

$$\Sigma = \begin{bmatrix}3 & 2 & 1.5 & 1\\ & 4 & -1 & 1.5\\ & & 4 & -0.5\\ & & & 3\end{bmatrix}, \qquad \Omega = \begin{bmatrix}0.3 & 0.25\\ & 1\end{bmatrix}. \tag{35}$$

In these parameter values, 32.4% of the variance in log-prices can be explained by the instruments, i.e., the instruments are relatively strong. The correlation between the demand and log-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 Σ. In this situation, we found that the 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.

We generate 50 datasets by first simulating 50 pairs of log-price and demand shocks and then calculating the implied log-prices and shares. We then apply both the Bayesian procedure and the GMM procedure to each dataset, setting H to 200. Table 7 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. In the IV case, there is less information available in the sense that only part of the variation in the continuous X variable can be used

Table 7
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
$\bar{\theta}_{\text{price}}$	0.28	10	0.33	−0.02
Σ_{11}	3.82	315.49	−1.59	6.86
Σ_{22}	3.11	383.2	−1.51	8.74
Σ_{33}	3.68	6301.31	−1.30	19.09
Σ_{44}	0.75	104.68	−0.06	4.02
Σ_{12}	2.33	117.63	−1.24	1.59
Σ_{13}	1.64	82.45	−1.00	1.20
Σ_{23}	1.92	139.48	0.78	2.65
Σ_{14}	0.36	38.25	−0.25	−1.42
Σ_{24}	0.56	24.03	−0.32	−1.05
Σ_{34}	0.20	24.87	0.10	−1.89
δ_1	0.002	0.002	0.003	0.001
δ_2	0.002	0.002	0.002	0.001
δ_3	0.002	0.002	−0.002	−0.003
δ_4	0.004	0.004	−0.005	−0.002
Ω_{11}	0.0002	0.0002	0.0012	−0.0003
Ω_{12}	0.002	0.003	−0.02	−0.01
Ω_{22}	0.03	1.28	−0.16	0.39
Corr_{Ω}	0.003	0.008	−0.002	−0.062

to infer about model parameters, holding all other aspects of the problem constant. It is, therefore, not surprising that the efficiency differential between the Bayes and GMM estimators widens. In some applications, the instruments used are notably weaker than in the simulation study considered here. It is well known that instrumental variable methods do not perform well under weak instrument conditions. In particular, there is a large literature devoted to obtaining improved asymptotic approximations for the purpose of inference. Bayesian methods can be particularly useful in the weak instrument case as they do not require asymptotic approximations of any kind. In the case of a linear structural equation, Conley et al. (2008) find that Bayesian methods provide improved sampling properties over IV and LIML methods for the weak instrument case. Our experience has borne out that when the information content of the sample declines the GMM estimator declines relative to the Bayes estimator. Therefore, it is logical to conclude that in the case of weak instruments we can expect that the advantage of the Bayes method relative to the GMM would remain.

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 are 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, there have been few alternatives to a GMM estimation approach.

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. In any event, those investigators interested in the elasticity of expected aggregate demand may require a distributional assumption regarding the demand shock. Whatever an investigator's position on the efficiency versus robustness debate, we argue that there should be interest in exploring this trade-off.

We conduct sampling experiments to demonstrate the efficiency advantages of a Bayesian approach as well as the performance of Bayes estimators for mis-specified models. For the range of specifications considered, Bayes estimates exhibit lower MSE with little bias. The Bayes estimator is not sensitive to the departures from normality, independence or homoskedasticity considered. GMM estimates are sensitive to the number of simulation draws used in computed expected share while Bayes estimators are insensitive to the number of draws. In addition, we find that the asymptotic standard errors computed for GMM estimates understate the true sampling variance by about a factor of 2.

Our Bayes estimator should be considered a practical alternative to GMM estimators for those who are willing to consider a somewhat less robust estimator but with what appears to be considerably greater efficiency.

Acknowledgements

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.

Appendix A. Prior settings

The hyperparameter values used in the sampling experiment and the empirical example are as follows:

$$\begin{aligned} &\overline{\theta_0} = \mathbf{0}, \qquad V_{\bar{\theta}} = 100 I_K \\ &\nu_0 = K+1, \qquad s_0^2 = 1 \end{aligned} \tag{A.1}$$

where I_K is an identity matrix of size K.

Recall that we reparameterize Σ as follows:

$$\Sigma = U'U$$
$$U = \begin{bmatrix} e^{r_{11}} & r_{12} & \cdots & r_{1K} \\ 0 & e^{r_{22}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{K-1,K} \\ 0 & \cdots & 0 & e^{r_{KK}} \end{bmatrix}. \tag{A.2}$$

First, note that the diagonal r's and the off-diagonal r's enter U in different scales: the former are exponentiated while the latter are not. This implies that we should allow the diagonal r's to have different priors from the off-diagonal r's.

Second, the jth diagonal element in Σ can be expressed as

$$\begin{aligned} &\Sigma_{jj} = \exp\left(2r_{jj}\right) + \sum_{m=1}^{j-1} \left(r_{mj}\right)^2 \quad j = 2, \ldots, K \\ &\Sigma_{11} = \exp\left(2r_{11}\right). \end{aligned} \tag{A.3}$$

If we simply let all the K diagonal r_{jj}'s follow the same prior distribution and all the off-diagonal r_{mj}'s follow another prior distribution, then the implied prior variance on Σ_{jj} will increase with j. An undesirable implication is that the order in which brands or products are organized could potentially make a difference.

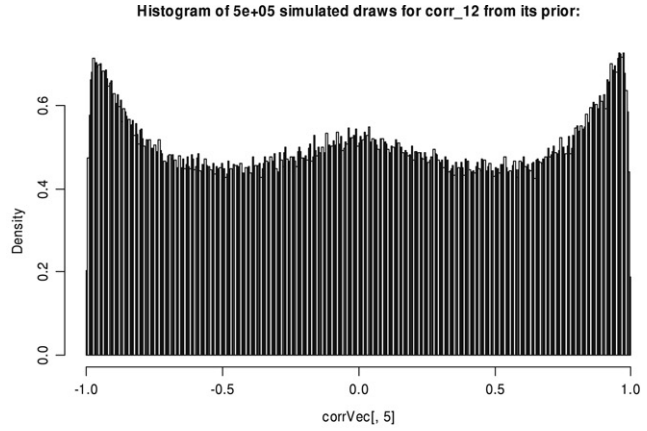
In order to achieve equal prior variances on Σ_{jj}, we allow the diagonal r_{jj}'s to have *different* prior distributions, $r_{jj} \sim N\left(0, \sigma^2_{r\_jj}\right)$, $j = 1, \ldots, K$, while the off-diagonal r_{mj}'s follow another prior distribution $N\left(0, \sigma^2_{r\_off}\right)$. In such setting, the implied prior variance of the diagonals in Σ is

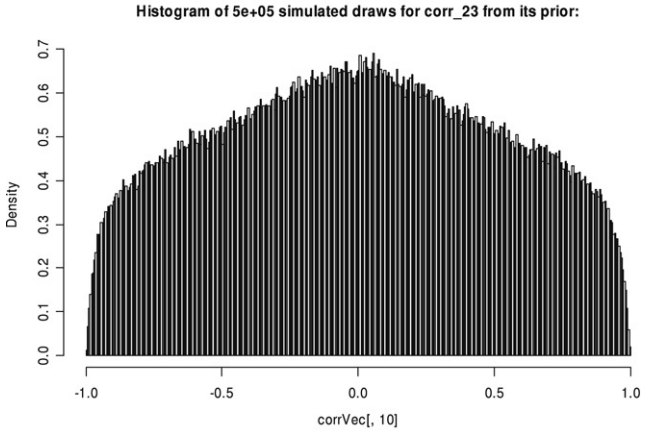
$$\begin{aligned} \text{var}\left[\Sigma_{jj}\right] = {} & 2(j-1)\sigma^4_{r\_off} + \exp\left(8\sigma^2_{r\_jj}\right) - \exp\left(4\sigma^2_{r\_jj}\right), \\ & j = 1, \ldots, K. \end{aligned} \tag{A.4}$$

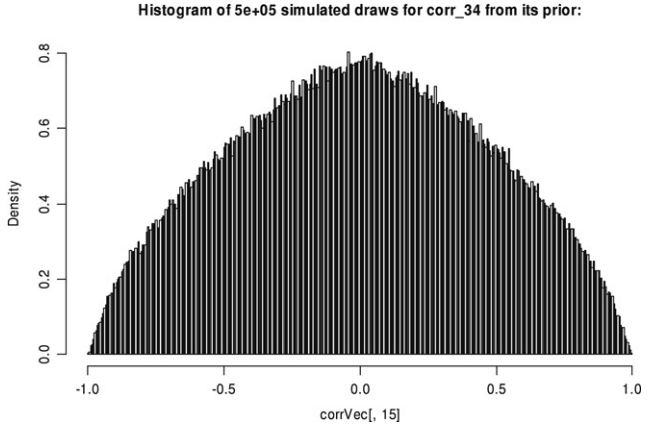
Third, denote the equal prior variance with $\text{var}\left[\Sigma_{jj}\right] = c$, for all j. We can solve for $\sigma^2_{r\_jj}$ once $\sigma^2_{r\_off}$ is assessed.

$$\sigma^2_{r\_jj} = \frac{1}{4}\log\left(\frac{1+\sqrt{1-4\left(2\left(j-1\right)\sigma^4_{r\_off}-c\right)}}{2}\right). \tag{A.5}$$

We choose $\sigma^2_{r\_off} = 1$ and $c = 50$ so that the elements of Σ have diffuse priors and that the priors of the associated correlations are roughly uniformly distributed between 0 and 1, as shown below.







These choices imply that $\sigma^2_{r\_jj}$ should be set to $\sigma^2_{r\_11} = 0.5067$, $\sigma^2_{r\_22} = 0.5019$, $\sigma^2_{r\_33} = 0.4970$, $\sigma^2_{r\_44} = 0.4918$. Note that these variances are very similar in magnitude. Thus, even if we did not allow for different priors on the diagonal r_{jj}'s, but, instead assessed the same prior variance for all diagonal r_{jj}'s, the induced prior on the diagonal elements of Σ would be roughly symmetric.

In the IV examples considered in Section 8, we keep the same priors on $\bar{\theta}$ and r as in the demand model. The hyperparameters for δ and Ω are

$$\bar{\delta} = \mathbf{0}, \quad V_\delta = 100 I_{J+1}, \quad \nu_0 = K+1, \quad V_\Omega = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}. \tag{A.6}$$

Appendix B. GMM procedure

The theoretical moment condition is $E\left[Z_t'\eta_t\right] = 0$, where Z_t is J (number of alternatives) by M (number of moments), and η_t is J by 1. The sample analogue is $\hat{m}_T\left(\bar{\theta}, \Sigma\right) = \frac{1}{T}\sum_{t=1}^{T} Z_t'\left(\hat{\mu}_t\left(\Sigma\right) - X_t\bar{\theta}\right)$.

The GMM objective is

$$g\left(\bar{\theta}, \Sigma\right) = \hat{m}_T\left(\bar{\theta}, \Sigma\right)' A^{-1} \hat{m}_T\left(\bar{\theta}, \Sigma\right) \tag{B.1}$$

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

If we take the 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 Σ and weighting matrix A:

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

where X is JT by K, and Z is JT by M. Thus, the GMM search is limited to Σ.

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, restart the optimization routine from the converged estimates to ensure that the GMM estimates converged. Use the final converged results $=> \hat{\Sigma}_{GMM}, \hat{\bar{\theta}}_{GMM}, \hat{\eta}_{jt}^{(2)}$.

GMM Standard Errors:

Standard results in the GMM literature provide an asymptotic approximation to the sample variance of the GMM estimator.

Let $\psi = \left(\bar{\theta}, r\right)$; then

$$\text{Var}\left(\hat{\psi}_{GMM}\right) = \frac{1}{T}\left(D'\hat{V}^{-1}D\right)^{-1}\Big|_{\psi=\hat{\psi}_{GMM}}. \tag{B.2}$$

D is a matrix of derivatives of the GMM criterion, $g(\cdot)$, w.r.t. to the parameters. $\hat{V}$ is a consistent estimate of the variance of $\hat{m}_T$.

$$D = \left[\frac{\partial m_T}{\partial \bar{\theta}} \quad \frac{\partial m_T}{\partial r}\right] \tag{B.3}$$

with

$$\frac{\partial m_T}{\partial \bar{\theta}} = -\frac{1}{T}\sum_t Z_t'X_t \tag{B.4}$$

and

$$\frac{\partial m_T}{\partial r} = \frac{1}{T}\sum_t Z_t'\left(\frac{\partial \hat{\mu}_t}{\partial r}\right) \tag{B.5}$$

$\hat{\mu}_t$ is the mean utility. The derivatives of the mean utility w.r.t. r are computed numerically.

References

Berry, Steven, 1994. Estimating discrete-choice models of product differentiation. RAND Journal of Economics 25 (2), 242–262.

Berry, Steven, Levinsohn, James, Pakes, Ariel, 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Chen, Yuxin, Yang, Sha, 2007. Estimating disaggregate models using aggregate data through augmentation of individual choice. Journal of Marketing Research 44, 613–621.
Chernozhukov, V., Hong, H., 2005. An MCMC approach to classical estimation. Journal of Econometrics 115, 148–1509.
Chintagunta, P., Dube, J.P., Singh, V., 2003. Balancing profitability and customer welfare in a supermarket chain. Quantitative Marketing and Economics 1, 111–147.
Conley, Tim, Hansen, Chris, McCulloch, Rob, Rossi, Peter, 2008. A semi-parametric Bayesian approach to the instrumental variables problem. Journal of Econometrics 144, 276–305.
Davis, Peter, 2006. Spatial competition in retail markets: Movie theaters. RAND Journal of Economics 37 (4), 964–982.
Gallant, A.R., Hong, H., 2007. A statistical inquiry into the plausibility of recursive utility. Journal of Financial Econometrics 5, 523–559.
Goettler, Ronald, Shachar, Ron, 2001. Spatial competition in the network television industry. RAND Journal of Economics 32 (4), 624–656.
Goldfarb, Avi, Lu, Qiang, Moorthy, Sridhar, 2005. Measuring brand equity in an equilibrium framework: A structural approach. Working Paper, 2005 QME Conference.
Musalem, Andres, Bradlow, Eric, Raju, Jagmohan, 2006. Bayesian estimation of random-coefficients choice models using aggregate data. Journal of Applied Econometrics (in press).
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, Allenby, Greg, McCulloch, Robert, 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, Sofia, 2004. Vertical contracts between manufacturers and retailers: Inference with limited data. CUDARE Working Paper, UC, Berkeley.
Yang, Sha, Chen, Yuxin, Allenby, Greg, 2003. Bayesian analysis of simultaneous demand and supply. Quantitative Marketing and Economics 1, 251–275.