Likelihood Based Estimation of Nonlinear Equilibrium Models with Random Coefficients

V. Brian Viard
Stanford Graduate School of Business, Stanford, CA 94304

Nicholas G. Polson
University of Chicago, Chicago IL 60637

Anne Gron
Northwestern University, Evanston, IL 60208.[1]

This draft: September 2002

Summary

In this paper we develop a likelihood based approach for estimating the equilibrium price and shares in markets with differentiated products and oligopoly supply. We model market demand using a discrete choice model with random coefficients and random utility. For most applications the likelihood function of equilibrium prices and shares is intractable and cannot be directly analyzed. To overcome this, we develop a Markov Chain Monte Carlo simulation strategy to estimate parameters and distributions based . To illustrate our methodology, we generate a dataset of prices and quantities simulated from a differentiated goods oligopoly across a number of markets. We apply our methodology to this dataset to demonstrate its attractive features as well as its accuracy and validity.

JEL Classification: D43, C33, C15, C11

Keywords: Market Equilibrium, Oligopoly, Differentiated Goods, MCMC.

[1]Anne Gron is Assistant Professor of Management and Strategy at the Kellogg School of Management, Northwestern University. Nicholas G. Polson is Professor of Econometrics and Statistics at Graduate School of Business, University of Chicago. V. Brian Viard is Assistant Professor of Strategic Management at the Graduate School of Business, Stanford University (viard_brian@gsb.stanford.edu). We gratefully acknowledge the comments of David Besanko, Lars Hansen, James Heckman and Alan D. Viard. We would like to thank the seminar participants at Ohio State University and the University of Chicago. Corresponding author: V. Brian Viard, 518 Memorial Way, Stanford Graduate School of Business, Stanford University, Stanford, CA 94305-5015, viard_brian@gsb.stanford.edu.

1 Introduction

In this paper, we develop a likelihood based method for estimating market equilibrium prices and shares for differentiated products produced by oligopolists. Our methodology extends the literature on estimating differentiated goods demand and market equilibrium by bringing together three existing approaches. One is the likelihood based estimation of equilibrium price and quantity in markets with differentiated goods as in Bresnahan (1981, 1987). His approach differs from ours in that the demand for differentiated products is based upon a simplified discrete choice model where products are ordered along one dimension, there are no random coefficients and there are no unobserved demand characteristics. In the second approach, researchers use likelihood based simulation methodologies to estimate random utility, random coefficient demand as in Rossi, McCulloch and Allenby (1996). In the third approach, researchers combine supply information with random coefficient, random utility demand and estimate equilibrium outcomes as in Berry, Levinsohn and Pakes (1995) using Generalized Method of Moments. In this paper, we develop a likelihood based estimation methodology for these models that uses a random utility, random coefficient demand model, oligopoly supply, and allows for unobserved demand and supply characteristics.

In general, the likelihood functions for the set of equilibrium prices and market shares arising from random coefficient demand with oligopoly supply results are computationally intractable. We propose an alternative solution to this problem by simulating the equilibrium posterior distribution of tastes and parameters from two conditional likelihoods, one based upon prices and the other upon shares. Our estimation methodology then simulates the equilibrium posterior distribution of tastes given the joint distribution of prices and shares. Basing estimation on these conditional likelihoods allows us to simulate the posterior distribution of parameters whose joint likelihood cannot be directly computed. We demonstrate the equivalence of these two methods in the simplified case of linear supply and demand where both the likelihood and the conditional likelihoods can be computed. Further, we estimate equilibrium prices and shares using our approach on a simulated dataset to illustrate the accuracy of the methodology. Our simulation approach uses Markov Chain Monte Carlo (MCMC henceforth) techniques, which are also used in the likelihood based estimation of random coefficient demand. Our specification relies on a logit-based demand system rather than the probit based structure used in demand-only likelihood approaches because the probit leads to an intractable market equilibrium system without individual level data.

Our likelihood based estimation methodology has a number of attractive features. It does not impose parametric assumptions upon the estimated distribution of consumer tastes. While the researcher does specify prior distributions for consumer tastes, the simulation methodology combines this prior information with the data and yields an estimated distribution of tastes from the simulation. Thus the prior might be normal, but the estimated distribution could be, for example, multi-modal. This is an attractive feature when consumers may be segmented by their preferences for certain features, such as brand, but the researcher specifies a unimodal distribution.[2] Another attractive feature is that standard errors can be calculated directly from the estimation data, whereas estimators using Generalized Methods of Moment require the researcher to calculate the appropriate weighting matrix to obtain standard errors. Theoretical convergence properties for our MCMC simulation are easily established (for example, Roberts and Smith, 1993) and simulation errors are well specified (Tierney, 1994). Functions of the estimated parameters of interest to researchers such as such as elasticities and variances are relatively straightforward to calculate. Another important difference between our approach and existing strategies for such estimation is that this technique can incorporate additional information in a flexible way. Researchers can use the model priors to combine information from other sources and to put more or less emphasis on this information. Finally, we note that when the likelihood is exact our estimation is also efficient.[3]

The rest of the paper proceeds as follows. Section 2 describes the demand and supply models that compose the market equilibrium. Section 3 discusses the issues involved with a likelihood based approach to estimating these models. We propose a method that is based on directly simulating the posterior distribution of tastes and parameters and show its equivalence to analysis of the likelihood in a simple linear equilibrium model. Section 4 discusses the specifics of implementing the likelihood approach. In section 5 we apply our estimation algorithm to a simulated dataset of differentiated goods and oligopoly supply. Our estimation algorithm provides reasonable estimates with significant improvement in precision over a simple logit-based demand approach. Section 6 concludes.

2 Random Utility-Random Coefficient Demand and Oligopoly Supply

Our demand model is based upon a random utility, random coefficient specification now commonplace in the analysis of differentiated goods demand. This model derives from the work of McFadden (1974) and can be applied to discrete choice problems and to differentiated products (see Anderson, De Palma and Thisse (1992), and Besanko, Perry and Spady (1990)).[4]

The demand structure arises from heterogeneous consumers choosing the product

[2]Besanko, Dube, and Gupta (2002) present an alternative method of incorporating multiple segments using a Generalized Method of Moments Estimator. Unlike their approach, our methodology does not require the researcher to specify the number of segments.

[3]An Efficient Method of Moments (EMM) (Gallant and Tauchen, 1996) approach could be used as well.

[4]Berry (1994) provides a summary of different models and their attributes and Nevo (2000) gives a brief summary of recent applications.

that maximizes their utility. Utility is a function of product attributes and disposable income. Goods' characteristics are both observable and unobservable to the econometrician. Consumers differ in their marginal utility of product attributes and in their idiosyncratic tastes for a particular product. Thus utility to consumer i from product j is:

$$u_{ij} = \alpha \log(y - p_j) + x_j\beta + \xi_j + \epsilon_{ij}$$

where $y - p_j$ is disposable income, x_j are observable product characteristics, ξ_j are unobservable product characteristics and ϵ_{ij} is an individual-product specific unobservable.[5] We assume that consumer tastes for characteristics, β, come from a distribution we denote as $p(\beta)$ with a mean of $\bar{\beta}$ and variance-covariance matrix $\Sigma_\beta = \sigma_\beta \sigma'_\beta$. The marginal utility of income, α, is also random with a mean of $\bar{\alpha}$ and variance σ^2_α. ξ_j is a normal random variable with mean zero and variance σ^2_ξ. The model relies on a logit demand structure, so the individual product-specific unobservable, ϵ_{ij} is drawn from a Type-1 extreme value distribution. Our specification includes income, y, which is generally not observable for market level data. In our data example this is constant, however it can be drawn from a distribution based upon population characteristics. For applications where income effects are negligible income is usually omitted.

Consumers purchase the good that gives them the highest utility. Let D_i denote consumers i's choice, then $D_i = l$ if $u_{il} > u_{ik}$ for all products $k \neq l$. Using the extreme value distribution of ϵ_{ij} the probability that a given individual purchases good l conditional on the parameters is multinomial logit (McFadden, 1974):

$$p\left(D_i = l | X, \alpha, \beta, y, \xi\right) = \frac{\exp\left(x'_l\beta + \alpha \log(y - p_l) + \xi_l\right)}{\sum_{j=0}^{J} \exp(x'_j\beta + \alpha \log(y - p_j) + \xi_j)}. \tag{1}$$

Since we will analyze markets in which individual transaction data are not available, we relate this probability to aggregate market shares. To do so we first note that the fraction of similar individuals in a large market (many consumers) who purchase good l will converge to the probability that a given individual in that market purchases good l, and assume that the probability of a tie in the utility ranking is zero. The market share of product l, denoted by s_l, is then determined by integrating over the distribution of tastes in the population and the distribution of unobserved characteristics over the set of products:

$$s_l(p, x) = \int_{\alpha,\beta,\xi} \frac{\exp\left(x'_l\beta + \alpha \log(y - p_l) + \xi_l\right)}{\sum_{j=0}^{J} \exp(x'_j\beta + \alpha \log(y - p_j) + \xi_j)} p(\alpha, \beta, \xi) d\alpha d\beta d\xi \tag{2}$$

where $p(\alpha, \beta, \xi)$ is the probability density function for (α, β, ξ). Equation (2) describes aggregate market demand for good l.

[5] While data will generally include observations across markets differentiated by time or by geography, for now we suppress the market subscript for clarity.

The supply side of the market is composed of a small number of multiproduct firms. A firm, f, offers a particular product j if j is a member of the set of products offered by f which we denote as Ω_f. Firm f's profit is then:

$$\pi_f = L \sum_{j \in \Omega_f} (p_j - mc_j) s_j(p, x)$$

where mc_j is the marginal cost of product j, L is total market size and $s_j(p, x)$ is market share of the product from consumers' demand. We specify marginal cost of product j as a function of an M-vector of observed cost characteristics, w_j, and unobserved (to the econometrician) factors, η_j:

$$mc_j = \exp(w_j \gamma + \eta_j)$$

We assume that, from the econometrician's perspective, η_j is a normal random variable with mean zero and variance σ_η^2. The M-column vector of parameters γ represent the contribution of each characteristic to marginal cost.

We assume that firms choose product prices to maximize profit, thus the first-order condition for product j is:

$$s_j(p, x) + (p_j - mc_j) \frac{\partial s_j(p, x)}{\partial p_j} + \sum_{k \in \Omega_f, k \neq j} (p_k - mc_k) \frac{\partial s_k(p, x)}{\partial p_j} = 0 \qquad (3)$$

where the derivatives are of (2) with respect to price. For the pricing derivative, we assume that the marginal utility of income, α, is constant.

The first-order condition can be transformed into:

$$(y - p_j) - \alpha(p_j - mc_j)(1 - s_j) + \alpha \sum_{k \in \Omega_f, k \neq j} (p_k - mc_k) s_k = 0.$$

Stacking the first-order conditions across products and rearranging we get:

$$\Lambda^{-1}(s)(p - mc) = -\alpha^{-1}(y - p)$$

where

$$\Lambda^{-1}(s) = \begin{pmatrix} (s_1 - 1) & H_{12}s_2 & \dots & H_{1J}s_J \\ H_{21}s_1 & (s_2 - 1) & \dots & H_{2J}s_J \\ \dots & \dots & \dots & \dots \\ H_{J1}s_1 & \dots & H_{JJ-1}s_{J-1} & (s_J - 1) \end{pmatrix}$$

$$p = \begin{pmatrix} p_1 \\ \dots \\ p_J \end{pmatrix}, \; mc = \begin{pmatrix} mc_1 \\ \dots \\ mc_J \end{pmatrix}, \; y - p = \begin{pmatrix} y - p_1 \\ \dots \\ y - p_J \end{pmatrix}.$$

This formulation separates the demand derivatives into a function of parameter α and a matrix, $\Lambda_t(s_t)$, which is independent of unknown parameters. H_t is an indicator variable which defines the association of products offered by each firm: $H_{jlt} = 1$ if products j and l are produced by the same firm and zero otherwise.

This equation can be re-written to isolate marginal cost. This yields:

$$p + \alpha^{-1}\Lambda(s)(y - p) = mc$$

or, after taking logarithms and substituting for marginal cost,

$$\log\left(p + \alpha^{-1}\Lambda(s)(y - p)\right) = w\gamma + \eta \tag{4}$$

where

$$w = \begin{pmatrix} w_1' \\ \dots \\ w_J' \end{pmatrix}, \eta = \begin{pmatrix} \eta_1 \\ \dots \\ \eta_J \end{pmatrix}.$$

We now turn to our likelihood based estimation procedure.

3 Likelihood Based Estimation

Ideally we would like to be able to derive an exact joint likelihood function for the parameters given shares and prices upon which to base our estimation. In theory we could determine the Jacobian implicitly defined by the joint system of (4)- the pricing equation- and an aggregate demand specification based upon (1) given a joint probability model for the unobserved characteristics. We could then determine the joint likelihood function for the parameters given shares and prices, $\mathcal{L}(s, p|\Theta)$ where $\Theta = (\alpha, \beta, \gamma)$. Next, we could combine the joint likelihood with the prior distribution of the parameters via Bayes' rule and directly generate the parameters of the system given the data.

In practice, however, the joint likelihood arising from a simultaneous equation system cannot generally be analyzed by simulation techniques unless the equations are linear. We propose an alternative methodology to analyze more complex simultaneous equation models. Rather than analyze the joint likelihood we analyze a model implicitly defined by two conditional distributions based upon our demand and supply models. We will refer to this as the conditional likelihood system. One can use the conditional likelihood system to simulate the posterior distribution of the parameters given the joint distribution of prices and shares. For the simple case of a linear system where we can also compute the actual likelihood, we show that the conditional likelihood system is equivalent to the actual likelihood discuss conditions under which the conditional likelihood will be closer to the actual likelihood for more general supply and demand systems.[6] Finally, we discuss the implementation of instrumental variables in our conditional likelihood system.

[6] To evaluate the validity of our approach in a more general setting in practice, we apply our methodology to a panel of simulated market equilibria in section 5.

Our problem is to find the distribution of tastes that would result in the observed market shares given random coefficient-random utility demand and oligopoly supply. Since we are developing a likelihood based approach, we do not use the aggregate, expected market share defined in (2) as the basis for estimation since this is a number and not a random variable. Instead, we use the share conditional on the parameters which is also the probability of purchase conditional on the parameters defined in (1) as a basis for our demand likelihood. As the share conditional on the parameters is the probability of purchase conditional on the parameters we have:

$$(s_l|\alpha,\beta,\xi) \approx p\left(D_i = l|\alpha,\beta,\xi\right) = \frac{\exp\left(x_l'\beta + \alpha\log(y-p_l) + \xi_l\right)}{\sum_{j=0}^{J}\exp(x_j'\beta + \alpha\log(y-p_j) + \xi_j)}.$$

Using (1) as the basis for the likelihood is discussed in Chamberlain (1999) and Rossi, McCulloch and Allenby (1996) with individual level data. We transform this expression into one that is linear in the utility terms by taking logs and subtracting the expected share of the outside good, applying the argument made in Berry (1994). Thus we have the probability of observing s_l conditional on the parameters, $p(s_l|\alpha,\beta,\xi)$ as:

$$p(s_l|\alpha,\beta,\xi) = \log(s_l|\alpha,\beta,\xi) = \log(s_0) + x_l'\beta + \alpha\log(y-p_l) + \xi_l. \tag{5}$$

This is also the probability that an individual with characteristics α and β will purchase a product with features x_l and ξ_l. We combine this with our priors about the distribution of tastes in order to generate our demand estimation. Our demand estimation essentially simulates "individuals" based upon the prior distribution of tastes, the prior distribution of the marginal utility of income and the data and then integrates them to generate a market share. Thus our demand estimation will calculate:

$$p(s_l) = \int_{\alpha,\beta} p(s_l|\alpha,\beta)p(\alpha,\beta)d\alpha d\beta \tag{6}$$

and share is no longer conditional on ξ since ξ is the error term in our estimation. We deal with the endogeneity of ξ by employing instrumental variables as presented in Section 3.3.

We now return to the problem of generating and estimating the likelihood associated with (4), the pricing equation, and (5), the demand equation. Since our estimation integrates over α and β we effectively estimate (4) and (6). Let $p(\xi,\eta|\Theta)$ be the joint probability distribution for the unobserved characteristics and let $\mathcal{J}(s,p:\Theta)$ be the Jacobian associated with (4) and (6). The joint likelihood function is:

$$\mathcal{L}(s,p|\Theta) = p(\xi(s,p),\eta(s,p)|\Theta)|\mathcal{J}(s,p:\Theta)|.$$

As we noted earlier, only very restrictive conditions such as linearity lead to a likelihood that has a closed-form solution. Thus solving directly for the joint likelihood is not possible in most situations. Moreover, the joint likelihood is generally not in a form where simulation techniques can be applied. Thus, we analyze the alternative model implicitly defined by the two conditional distributions based upon equations (4) and (6). This alternative approach is equivalent to analyzing the joint likelihood directly if it were available in the sense that the posterior distribution of the simulated parameters is the equilibrium distribution of the parameters given the joint distribution of the prices and shares. Subject to mild regularity conditions, these conditional distributions define a unique joint distribution $p\left(s,p|\Theta\right)$ by the Hammersley-Clifford Theorem (Besag, 1974).[7] The conditional likelihoods allow us to analyze nonlinear simultaneous systems and to incorporate additional error from ψ and η. The unique posterior distribution of the parameters will then be determined by Bayes' Rule via application of MCMC. In the case of a linear system, we can demonstrate the equivalence of these two approaches directly as in section 3.1.

Our conditional likelihood system is defined by:

$$\log(s)|p,\Theta \sim \mathcal{N}\left(g(p,y,\alpha,\beta),\sigma_\xi^2 I_J\right) \tag{7}$$

$$h(p,s,y,\alpha)|s,\Theta \sim \mathcal{N}\left(w\gamma,\sigma_\eta^2 I_J\right) \tag{8}$$

where:[8]

$$\begin{aligned} g\left(p,y,\alpha,\beta\right) &= \log\left(s_0\right) + X\beta + \alpha\log\left(y-p\right) + \xi \\ h\left(p,s,y,\alpha\right) &= \log\left(p + \alpha^{-1}\Lambda(s)(y-p)\right). \end{aligned}$$

The demand-side conditional likelihood is then proportional to:

$$\left(\sigma_\xi^2\right)^{-\frac{JT}{2}} \exp\left(-\frac{1}{2\sigma_\xi^2}||\log(s) - g(p,y,\alpha,\beta)||^2\right)$$

and the supply-side conditional likelihood is proportional to:

[7] The Hammersley-Clifford Theorem provides conditions under which a set of conditional distributions characterizes a unique joint distribution. It is based on the Besag formula which states that for any pair $(s^{(0)},p^{(0)})$ of points, the joint density $p(s,p|\Theta)$ can be determined as

$$\frac{p(s,p|\Theta)}{p(s^{(0)},p^{(0)}|\Theta)} = \frac{p(s|p^{(0)},\Theta)p(p|s,\Theta)}{p(s^{(0)}|p^{(0)},\Theta)p(p^{(0)}|s,\Theta)}$$

as long as a positivity condition is satisfied. The positivity condition in our case requires that for each point in the sample space, $p(s,p|\Theta)$ and the marginals have positive mass.

[8] We have stacked the variables here as:

$$X = \begin{pmatrix} x_1' \\ \dots \\ x_J' \end{pmatrix},\ y-p = \begin{pmatrix} y-p_1 \\ \dots \\ y-p_J \end{pmatrix},\ \xi = \begin{pmatrix} \xi_1 \\ \dots \\ \xi_J \end{pmatrix}.$$

$$\left(\sigma_\eta^2\right)^{-\frac{JT}{2}} \exp\left(-\frac{1}{2\sigma_\eta^2}||h(p,s,y,\alpha) - w\gamma||^2\right).$$

where $||x||^2 = x'x$ and T is the number of markets served.

We denote the joint likelihood defined by the conditional distributions by $\tilde{\mathcal{L}}(s,p|\Theta)$, which can be computed as:

$$\tilde{\mathcal{L}}(s,p|\Theta) = \int_{\xi,\eta} p(s,p|\Theta,\xi,\eta)p(\xi,\eta)d\xi d\eta.$$

In this case, we can proceed as if the distribution of $\log(s)$ conditional on p and parameters is a JT dimensional multivariate normal, with a mean of $g(p,y,\alpha,\beta)$ and variance-covariance matrix $\sigma_\xi^2 I$; and the distribution of $h(p,s,y,\alpha)$ is a JT dimensional multivariate normal with mean of $w\gamma$ and variance-covariance matrix $\sigma_\eta^2 I$. We can then simulate the equilibrium distribution.

MCMC methods provide a natural approach to solving this system. These methods break the problem up into a number of simpler problems. In principle, we combine the likelihood $\mathcal{L}(s,p|\Theta)$ with the random effects distribution of the parameters, $p(\Theta)$. Bayes' rule yields the posterior distribution of the parameters given the data:

$$p(\Theta|s,p) \propto \mathcal{L}(s,p|\Theta)p(\Theta).$$

It is this posterior distribution that MCMC simulates to perform likelihood based inference.

The inputs to develop an MCMC algorithm are the conditional likelihood functions and the distribution of tastes. Using these we calculate the distribution of each parameter conditional on the data and the other parameters, which we refer to as the parameter's conditional posterior distribution. We specify a sampling algorithm based upon combining two techniques known as the Gibbs sampler and the Metropolis-Hastings algorithm (Hastings, 1970). Our estimation process simulates iteratively from the conditional posterior distributions to define a Markov chain whose equilibrium distribution is the full joint posterior distribution. We then sample from the joint posterior distribution and estimate functions of interest by Monte Carlo averaging.

3.1 Illustrative Example: A Linear System

To demonstrate the equivalence of the conditional likelihood system to the actual likelihood, we now consider a simple, linear system. This stylized example of our general approach allows us to solve the likelihood $\mathcal{L}(s,p|\Theta)$ analytically and compare that to the joint likelihood $\tilde{\mathcal{L}}(s,p|\tilde{\Theta})$ derived from the conditional distributions. The linear equilibrium model has demand specified as $s = p\beta_i + \xi$ with $\beta_i \sim \mathcal{N}\left(\bar{\beta}, \sigma_\beta^2\right)$, $\xi \sim N(0, \sigma_\xi^2)$. Supply is given by $p = s\gamma_l + \eta$ and $\gamma_l \sim \mathcal{N}(\bar{\gamma}, \sigma_\gamma^2)$ with $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$. [9]

This system is equivalent to:

$$s = p\bar{\beta} + \epsilon_s \text{ and } p = s\bar{\gamma} + \epsilon_p$$

where $\epsilon_s \sim \mathcal{N}(0, \sigma_s^2)$, $\epsilon_p \sim \mathcal{N}(0, \sigma_p^2)$ $\sigma_s^2 = \sigma_\beta^2 pp' + \sigma_\xi^2$ and $\sigma_p^2 = \sigma_\gamma^2 ss' + \sigma_\eta^2$. Solving this equilibrium leads to a joint distribution for shares and prices which defines $\mathcal{L}(s,p|\Theta)$ where $\Theta = \{\bar{\beta}, \bar{\gamma}, \sigma_\beta^2, \sigma_\gamma^2\}$. More specifically,

$$s, p|\Theta \sim \mathcal{N}\left(0, \begin{pmatrix} \bar{\sigma}_s^2 & \rho\bar{\sigma}_s\bar{\sigma}_p \\ \rho\bar{\sigma}_s\bar{\sigma}_p & \bar{\sigma}_p^2 \end{pmatrix}\right)$$

where $\bar{\sigma}_s^2 = \frac{1}{1-\bar{\beta}\bar{\gamma}}[\sigma_\xi^2 + \bar{\beta}^2\sigma_\eta^2]$, $\bar{\sigma}_p^2 = \frac{1}{1-\bar{\beta}\bar{\gamma}}[\bar{\gamma}^2\sigma_\xi^2 + \sigma_\eta^2]$ and $\rho\bar{\sigma}_s\bar{\sigma}_p = \frac{1}{1-\bar{\beta}\bar{\gamma}}[\bar{\gamma}\sigma_\xi^2 + \bar{\beta}\sigma_\eta^2]$. The corresponding conditional distributions are given by:

$$s|p, \Theta \sim \mathcal{N}\left(\frac{\rho\bar{\sigma}_s}{\bar{\sigma}_p}p, \bar{\sigma}_s^2(1-\rho^2)\right) \text{ and } p|s, \Theta \sim \mathcal{N}\left(\frac{\rho\bar{\sigma}_p}{\bar{\sigma}_s}s, \bar{\sigma}_p^2(1-\rho^2)\right).$$

Under our methodology, the corresponding likelihood $\tilde{\mathcal{L}}(s,p|\tilde{\Theta})$ is defined by the conditional distributions:

$$s|p, \tilde{\Theta} \sim \mathcal{N}\left(\tilde{\beta}p, \tilde{\sigma}_s^2\right) \text{ and } p|s, \tilde{\Theta} \sim \mathcal{N}\left(\tilde{\gamma}s, \tilde{\sigma}_p^2\right)$$

where $\tilde{\Theta} = \{\tilde{\beta}, \tilde{\gamma}, \tilde{\sigma}_\xi^2, \tilde{\sigma}_\eta^2\}$. Therefore, the likelihood from the market equilibrium based on conditional densities, $\tilde{\mathcal{L}}(s,p|\tilde{\Theta})$ can be related to $\mathcal{L}(s,p|\Theta)$ by matching $\tilde{\beta} = \frac{\rho\bar{\sigma}_s}{\bar{\sigma}_p}$, $\tilde{\gamma} = \frac{\rho\bar{\sigma}_p}{\bar{\sigma}_s}$, $\tilde{\sigma}_s^2 = \bar{\sigma}_s{}^2(1-\rho^2)$ and $\tilde{\sigma}_p^2 = \bar{\sigma}_p{}^2(1-\rho^2)$. This relationship holds as long as the original model is identified (i.e. $\bar{\beta}\bar{\gamma} \neq 1$).

For a nonlinear system this line of reasoning does not directly apply. However, if the errors σ_ξ^2 and σ_η^2 are small, the system approximates a linear system locally (near the equilibrium) and the above argument applies.

In practice, identification of the system usually requires the use of instrumental variables beyond the identification provided by distributional and functional form assumptions. While we discuss instrumental variables in Section 3.3, we mention them here to note how their use relates to the equivalence argument. It is typical to assume that there are instruments z, v which satisfy the moment restrictions $E[z\xi] = 0$ and $E[v\eta] = 0$ and therefore the error variances $\sigma_\xi^2 zz'$ and $\sigma_\eta^2 vv'$ are extremely small. In practice then, we apply our methodology on the instrumented system with small σ_ξ^2 and σ_η^2. We also perform a sensitivity analysis of the resulting posterior distribution to the size of these errors.

[9]For illustrative purposes the parameters in the supply equation are random coefficients as well. Our equivalence result also holds when the supply does not have random coefficients.

Assume, for now, that the unobserved product characteristic, ξ, is normally distributed with mean zero and variance σ_ξ^2. Since β and ξ are independent normals, the vector of implied shares given by (5) can be rewritten as:

$$\log(s_t) = \alpha \log(y_t - p_t) + X_t\bar{\beta} + e_t^D \tag{9}$$

where

$$e_t^D = \begin{pmatrix} e_{1t}^D \\ \cdots \\ e_{Jt}^D \end{pmatrix},$$

$e_t^D \sim \mathcal{N}(0, X_t\Sigma_\beta X_t' + \sigma_\xi^2 I)$ and we have re-introduced the time subscripts. We have now subsumed the effect of the outside good in the intercept term and the $(1,1)$ element of Σ_β is $x_{0t}\Sigma_\beta x_{0t}'$.[10] So the observed vector of market shares is a vector of random variables from the perspective of the econometrician by virtue of the unobservables and random coefficients.

Similarly, the observed vector of prices of products in time period t, p_t, is a vector of random variables implicitly defined by:[11]

$$\log[p_t + \alpha^{-1}\Lambda_t(s_t)(y_t - p_t)] = w_t\gamma + e_t^S \tag{10}$$

where

$$e_t^S = \begin{pmatrix} e_{1t}^S \\ \cdots \\ e_{Jt}^S \end{pmatrix}$$

and $e_t^S \sim \mathcal{N}(0, \sigma_\eta^2)$. Λ_t is a function of the observed shares and the demand side identifies these observed shares as a function of the random variables. Similarly, shares are a function of observed prices in the demand equation and the supply side identifies these as a function of the random variables. This is the simultaneity in the system. As written, it appears that prices (defined by the supply side) are not functions of $\bar{\beta}$. However, they are because the shares defined by the logit demands are part of the Λ matrix. This can be most easily seen by referring back to our linear example. There it was clear that the supply-side depended on the demand-side parameters - they entered through

[10]As written here the first β coefficient is time-dependent. In our simulation in the next section we do not allow the characteristics of the outside good to vary across time periods and we use the same distribution of income in each time period so that there is no time variation in this coefficient. In an application with real data these assumptions would not hold and one would need to include time dummies in x_{jt}.

[11]Note that implicit in this representation is our assumption that α in the supply equation is not random, otherwise the variance component would contain a term depending on the variance of α.

the Jacobian. Similarly, in this nonlinear case, the supply side depends on the demand-side parameters through the Jacobian implied by the joint system. We now turn to the problem of incorporating instrumental variables.

3.3 Incorporating Instrumental Variables

Up to this point we have included no additional information to help identify the unobserved characteristics, ξ and η. In many cases the researcher has additional variables which may be used to help identify ξ and η. Suppose that the researcher has an R column-vector of demand-side instrumental variables, z_{jt}, and a Q column-vector of supply-side instrumental variables, v_{jt}, for each product j and time period t that can be used to identify the unobserved characteristics. The instruments satisfy the conditions $E[z_{jt}\xi_j] = 0$ and $E[v_{jt}\eta_t] = 0$. We impose these latter conditions through distributional assumptions: $z_t\xi \sim \mathcal{N}(0, \sigma_\xi^2 z_t z_t')$ and $v_t\eta \sim \mathcal{N}(0, \sigma_\eta^2 v_t v_t')$ where we set the variances to be very small and perform sensitivity analysis of the posterior to these variances. We can then instrument the system (9) and (10) by multiplying the demand side through by $z_t = [z_{1t} \ldots z_{Jt}]$ and the supply-side by $v_t = [v_{1t} \ldots v_{Jt}]$ to get:

$$z_t \log(s_t) = \alpha z_t \log(y_t - p_t) + z_t X_t \bar{\beta} + z_t e_t^D$$

$$v_t \log[p_t + \alpha^{-1}\Lambda_t(s_t)(y_t - p_t)] = v_t w_t \gamma + v_t e_t^S.$$

The problem now is to solve the simultaneous equilibrium of supply and demand specified by these two equations. These equations can be interpreted as conditional likelihood functions:

$$z_t \log(s_t)|v_t \log(A_t), \alpha, \bar{\beta}, \Sigma_\beta, y_t \sim \mathcal{N}(\alpha z_t \log(y_t - p_t) + z_t X_t \bar{\beta}, \Sigma_D) \quad (11)$$

$$v_t \log(A_t)|z_t \log(s_t), \alpha, \gamma, y_t \sim \mathcal{N}(v_t w_t \gamma, \Sigma_S)$$

where $\Sigma_D = (z_t X_t)\Sigma_\beta(z_t X_t)' + \sigma_\xi^2 z_t z_t'$, $\Sigma_S = \sigma_\eta^2 v_t v_t'$ and $A_t(p_t, \Lambda_t(s_t), \alpha) = [p_t + \alpha^{-1}\Lambda_t(s_t)(y_t - p_t)]$. We now turn to MCMC implementation.

4 MCMC Implementation

In order to estimate a model using our MCMC algorithm the researcher must first develop the conditional likelihood distributions for the equations of the simultaneous system model as we have done in Section 3. Next the researcher needs to determine which parameters can be grouped together for estimation and choose appropriate prior distributions. We do this in Section 4.1. Given the first two steps, one can then calculate the conditional posterior distributions using the conditional likelihood distributions and the prior distributions. This may require using latent variables for some parameters. Next, one applies the relevant simulation technique for each block of parameters, determines starting values and specifies the level of the "error" in the estimated system. The conditional posterior distributions, appropriate simulation methods, and starting values for our application are described in Section 4.2. These steps comprise the setup and the actual estimation procedure. Finally, the researcher will want to evaluate the estimation results by computing parameter estimates, taste distributions, and functions of the parameters such as elasticities and by evaluating the sensitivity of the estimates to particular estimation choices such as the level of "error." These issues are covered in section 4.3.

4.1 Specification of the Distribution of Tastes

We now discuss the issues in blocking the parameters and choosing prior distributions for the parameters. The distributional forms and blocks for parameters are chosen to balance analytical convenience, speed of the simulator and generality of the implied assumptions. MCMC simulators draw much faster from recognizable distributions (due to the lower rejection rate) so we emphasize choosing priors which lead to recognizable posterior conditional distributions. We also block together parameters that can be jointly drawn from the same recognizable posterior conditional distribution to speed up convergence. Blocking parameters involves balancing realistic priors for the parameters with the increased speed of the simulator that comes from reducing the number of blocks to be drawn. For example, variance parameters require a prior with only positive support while linear parameters can often be drawn from distributions whose support is the real line. We provide for generality in our priors primarily by employing diffuse priors so that prior information has little influence on the posteriors. Here it is convenient to block the parameters in as large groups as possible and simulate them together as described in Liu, Wong and Kong (1994).

We divide our parameters Θ into five blocks: $\bar{\beta}$, Σ_β, γ, α and σ_α^2 and assume independent priors so that:

$$p(\Theta) = p(\bar{\beta})p(\Sigma_\beta)p(\gamma)p(\alpha)p(\sigma_\alpha^2).$$

Our reasoning is as follows. The first three parameter blocks can all be drawn from recognizable posterior distributions as shown below. No pair of these blocks can be combined without sacrificing this recognizability. The linear parameters from the demand side cannot be combined with those from the supply side since they come from different conditional likelihoods. The variance and linear parameters from the demand-side cannot be combined because the priors for the former require positive support while those for the latter allow any real value. The two α blocks cannot be combined with other priors because these parameters enter in a nonlinear fashion and must be drawn from an unrecognizable distribution.

We specify the following prior distributions for each of the parameter blocks:

$$\alpha \sim \mathcal{N}(a_0, A_0^{-1}) I_{\alpha>0} \;\; \sigma_\alpha^2 \sim \mathcal{W}(a_2, A_2^{-1})$$
$$\bar{\beta} \sim \mathcal{N}(b_0, B_0^{-1}) \;\; \Sigma_\beta \sim \mathcal{W}(b_2, B_2^{-1})$$
$$\gamma \sim \mathcal{N}(g_0, G_0^{-1})$$

where $\mathcal{W}$ denotes the inverted Wishart distribution.[12] We use diffuse prior distributions for all the distributions as described below in our simulation results.

An alternative way to write our model is to use a three-stage hierarchical model which specifies the likelihood function in the first stage, the distributions of the parameters in the second stage and the priors in the third stage:

$$z_t \log(s_t)|p_t, \Theta \sim \mathcal{N}(\alpha z_t \log(y_t - p_t) + z_t x_t \beta, z_t z_t' \sigma_\xi^2)$$
$$\beta \sim \mathcal{N}(\bar{\beta}, \Sigma_\beta) \;\; \alpha \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2) I_{\alpha>0}$$
$$\bar{\beta} \sim \mathcal{N}(b_0, B_0^{-1}) \;\; \Sigma_\beta \sim \mathcal{W}(b_2, B_2^{-1}) \;\; \sigma_\alpha^2 \sim \mathcal{W}(a_2, A_2^{-1}).$$

The model for the supply-side contains only two stages since the coefficients are fixed. If the supply-side included random coefficients the second stage of the supply-side would specify the distributions for the coefficients:

$$v_t \log[p_t + \alpha^{-1}\Lambda_t(s_t)(y_t - p_t)]|s_t, \Theta \sim \mathcal{N}(v_t w_t \gamma, v_t v_t' \sigma_\eta^2)$$
$$\gamma \sim \mathcal{N}(g_0, G_0^{-1}).$$

We now turn to the MCMC algorithm for simulating from the posterior distribution.

4.2 Conditional Posterior Distributions

Before we derive the conditional posterior distributions for the parameters, it is necessary to discuss how we treat the parameters $\bar{\alpha}$, σ_α^2 and Σ_β. Due to the nonlinearity in the system, these distributions cannot be marginalized out directly. Hence we have to use latent variables or a Metropolis-Hastings step in our algorithm.

The likelihood for the parameters $\Theta = (\bar{\alpha}, \bar{\beta}, \sigma_\alpha^2, \Sigma_\beta, \gamma)$ from the full panel is:[13]

$$\mathcal{L}(s, p|\Theta) = \prod_{t=1}^{T} p(\log(s_t)|p_t, \Theta) p(p_t|\Theta). \tag{12}$$

[12]The inverted Wishart distribution is particularly convenient for specifying precision matrices because its posterior is positive-definite and also Wishart. The first parameter is the degrees of freedom and the second parameter is the variance-covariance matrix. Higher degrees of freedom place more mass of the distribution near zero while lower degrees of freedom spread the mass over the positive real line.

[13]We make a simplifying assumption here: $p^*(\Theta) = \frac{p(t)|\Theta)}{p(p(t)|log(s_t),\Theta)} \propto 1$. Specifically, we only have to assume that $p^*(\alpha)$ is flat in the region where the data is most informative about α as in the principle of stable estimation (Edwards, Lindman and Savage, 1963). We can then proceed as if the distribution $p^*(\alpha)$ is diffuse and close to uniform. We then have: $p(log(s_t), p_t|\Theta) = p(log(s_t)|p_t)p(p_t|\Theta)$.

To compute the likelihood we introduce vectors of latent parameters $\underline{\alpha} = (\alpha_1, \ldots, \alpha_T)$ and $\underline{\beta} = (\beta_1, \ldots, \beta_T)$ where $\alpha_t \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2) I_{\alpha_t > 0}$ and $\beta_t \sim \mathcal{N}(\bar{\beta}, \Sigma_\beta)$. This allows us to write the likelihood defined by (12) above as the marginal:

$$\mathcal{L}(s, p|\Theta) = \int_{\alpha_t, \beta_t} \prod_{t=1}^{T} p(\log(s_t)|\alpha_t, \beta_t, p_t) p(p_t|\alpha_t, \beta_t) p(\alpha_t, \beta_t|\Theta) d\alpha_t d\beta_t.$$

Bayes rule allows us to write the posterior $p(\Theta|s, p)$ as a marginal from the joint posterior $p(\Theta, \underline{\alpha}, \underline{\beta}|s, p)$. Hence we need only create a Markov chain that provides samples from the joint distribution $\{\Theta^{(g)}\}_{g=1}^{G}$. We now describe how to draw from the posterior conditional distributions to obtain draws from the joint posterior distribution. We divide the procedure into steps by the blocks described in the previous section.

1. Simulate from $\underline{\alpha}|\Theta, s, p$. As the α_t's are conditionally independent we can do this component-by-component and generate $\alpha_t|\Theta, s_t, p_t$ using a Metropolis algorithm. For notational simplicity let $\pi(\alpha_t) = p(\alpha_t|\Theta, s_t, p_t)$. Now:

$$\pi(\alpha_t) \propto p(\log(s_t)|\bar{\beta}, \Sigma_\beta, \alpha_t, p_t) p(p_t|\gamma, \alpha_t) p(\alpha_t|\bar{\alpha}, \sigma_\alpha^2)$$

and $p(\alpha_t|\bar{\alpha}, \sigma_\alpha^2) \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2) I_{\alpha \in \mathcal{A}}$ where $\mathcal{A} = \{\alpha_t{:}\alpha_t > \max(\psi_t)\}$ where ψ_t is defined below. Due to the normality and linearity in α_t of the demand side equation we have that $p(\log(s_t)|\bar{\beta}, \Sigma_\beta, \alpha_t, p_t) \propto \mathcal{N}(a_D, A_D)$ where:

$$A_D = (z_t \log(y_t - p_t))' (\Sigma_D)^{-1} (z_t \log(y_t - p_t))$$

$$a_D = A_D^{-1}[(z_t \log(y_t - p_t))' (\Sigma_D)^{-1} (z_t \log(s_t) - z_t x_t \bar{\beta})].$$

The supply side does not simplify due to the nonlinearity in α. For the Metropolis blanket we use an expansion of $\log(A_t)$ and use the approximation $\log(A_t) \approx \log(p_t) + \alpha_t^{-1}\psi_t$ where $\psi_t = \Lambda_t(s_t)(y_t - p_t)/p_t$ with element-by-element division. For the supply-side we have that $p(\log(p_t)|\gamma, \alpha_t) \propto \mathcal{N}(a_S, A_S)$ where:

$$\bar{A}_S = (v_t\psi_t)'(\Sigma_S)^{-1}(v_t\psi_t)$$

$$a_S = [\bar{A}_S^{-1}[(v_t\psi_t)'(\Sigma_S)^{-1}(v_t w_t \gamma - v_t \log(p_t))]]^{-1}$$
$$A_S = \bar{A}_S a_S^{-4}.$$

We combine this with the demand-side normal and prior to find a Metropolis blanket $Q(\alpha_t)$ for generating the new α_t. Note that we attempt to make our blanket approximate the posterior distribution as closely as possible so that candidate draws are concentrated in the mass of the distribution thereby increasing the acceptance rate and speed of convergence. Then, the Metropolis acceptance probability is used to correct the approximation so we finally have a draw from the correct equilibrium distribution. Hence our Metropolis blanket $Q(\alpha_t)$ combines these three normal distributions and is given by $Q(\alpha_t) \sim \mathcal{N}(a_3, A_3)$ where:

$$A_3 = A_S + A_D + \sigma^2_\alpha$$

$$a_3 = A_3^{-1}(A_D a_D + A_S a_S + \sigma^2_\alpha \bar{\alpha}).$$

If the current value of α_t in our algorithm is given by $\alpha_t^{(g)}$ and the next candidate draw is given by $\alpha_t^{(g+1)}$ drawn from $Q(\alpha_t^{(g+1)})$ then we accept this new draw with probability given by:

$$\min \left(\frac{\pi(\alpha_t^{(g+1)})/Q(\alpha_t^{(g+1)})}{\pi(\alpha_t^{(g)})/Q(\alpha_t^{(g)})} \right)$$

where $\pi(\alpha_t)$ is defined above.

2. Simulate from the conditional posteriors $\bar{\alpha}|\underline{\alpha}, \sigma^2_\alpha, s, p$ and $\sigma^2_\alpha|\underline{\alpha}, \bar{\alpha}, s, p$. Standard conjugate theory (see DeGroot, 1970) implies:

$$\bar{\alpha}|\underline{\alpha}, \sigma^2_\alpha, s, p \sim \mathcal{N}(a_1, A_1^{-1})$$

where

$$A_1 = T\sigma^2_\alpha + A_0$$

$$a_1 = A_1^{-1} \left(\sigma^2_\alpha \sum_{t=1}^{T} \alpha_t + A_0 a_0 \right)$$

and

$$\sigma^2_\alpha|\underline{\alpha}, \bar{\alpha}, s, p \sim Ga \left(\frac{T}{2} + a_2, (A_2^{-1} + \frac{1}{2} \sum_{t=1}^{T} (\alpha_t - \bar{\alpha})^2)^{-1} \right)$$

where Ga denotes a Gamma distribution.

3. Simulate the latent variables $\underline{\beta}|\bar{\beta}, \Sigma_\beta, \underline{\alpha}, s, p$. As the elements of the vector $\underline{\beta}$ are conditionally independent we can generate them component-by-component. The conditional posterior of β_t is given by:

$$\beta_t|\bar{\beta}, \Sigma_\beta, \underline{\alpha}, s, p \sim \mathcal{N}(b_3, B_3^{-1})$$

where

$$B_3 = (z_t x_t)' \Sigma_D (z_t x_t)$$

$$\hat{b}_3 = B_3^{-1}(z_t x_t)' B_3^{-1} \left(z_t \log(s_t) - \alpha_t z_t \log(y_t - p_t - z_t x_t \bar{\beta}) \right)$$

and

$$b_3 = (B_3 + \Sigma_\beta)^{-1} \left(B_3 \hat{b}_3 + \Sigma_\beta \bar{\beta} \right).$$

4. Simulate from the conditional posteriors $\bar{\beta}|\Sigma_\beta, \underline{\alpha}, s, p$ and $\Sigma_\beta|\bar{\beta}, \underline{\beta}, \underline{\alpha}, s, p$. From standard conjugate theory:

$$\bar{\beta}|\Sigma_\beta, \underline{\alpha}, s, p \sim \mathcal{N}(b_1, B_1^{-1})$$

where

$$B_1 = (zx)' R_\beta^{-1}(zx) + B_0$$

$$b_1 = B_1^{-1}\left((zx)' R_\beta^{-1}(z\log(S) - \alpha z \log(Y-P)) + B_0 b_0\right)$$

and

$$R_\beta = \sigma_\xi^2 \text{diag}[z_1 z_1', \ldots, z_T z_T'] + \text{diag}[(z_1 x_1)\Sigma_\beta^{-1}(z_1 x_1)', ..., (z_T x_T)\Sigma_\beta (z_T x_T)']$$

$$zx = \begin{pmatrix} z_1 x_1 \\ \ldots \\ z_T x_T \end{pmatrix}, z\log(Y-P) = \begin{pmatrix} z_1 \log(y_{i1} - p_1) \\ \ldots \\ z_T \log(y_T - p_T) \end{pmatrix}, z\log(S) = \begin{pmatrix} z_1 \log(s_1) \\ \ldots \\ z_T \log(s_T) \end{pmatrix}.$$

We restrict the taste parameters to be uncorrelated but allow for heteroskedasticity. The posterior conditional distribution of Σ_β is therefore given by:

$$\sigma_{\beta,k}^2|\underline{\beta}, \bar{\beta} \sim Ga\left(\frac{T}{2} + b_{2,k}, B_{2,k}^{-1} + \frac{1}{2}\sum_{t=1}^{T}(\beta_{t,k} - \bar{\beta}_k)^2\right)$$

where $\sigma_{\beta,k}^2$ is the (k,k) component of Σ_β, $b_{2,k}$ is the kth component of the prior mean, $B_{2,k}$ is the (k,k) component of the prior variance, $\beta_{t,k}$ is the kth component of the latent variable and $\bar{\beta}_k$ is the kth element of $\bar{\beta}$.

5. Simulate $\gamma|\underline{\alpha}, s, p$. The posterior conditional distribution follows from standard conjugate theory:

$$\gamma|\underline{\alpha}, s, p \sim \mathcal{N}(g, G)$$

where

$$G = (vw)' R_\gamma' (vw) + G_0$$

$$g = G^{-1}[(vw)' R_\gamma' v \log(A) + G_0 g_0]$$

where

$$R_\gamma = \sigma_\eta^2 \text{diag}\,(v_1 v_1', \ldots, v_T v_T')$$

and

$$vw = \begin{pmatrix} v_1 w_1 \\ \ldots \\ v_T w_T \end{pmatrix}, v\log(A) = \begin{pmatrix} v_1 \log(A_1) \\ \ldots \\ v_T \log(A_T) \end{pmatrix}.$$

These five conditional distributions specify our algorithm.

Given initial starting values for the parameters, we iterate through these steps and produce a sequence of G draws after discarding a burn-in period to reduce the dependence on the initial starting values. There are a number of convergence issues. First, a natural candidate for the initial starting point is just a logit regression for the demand side and a supply side regression given the α estimated from the logit regression. Of course these estimates are incorrect but they serve as a reasonable starting point. Mengersen and Robert (1998) provide a review of the general methodologies for assessing convergence from the MCMC output. From a theoretical perspective, as steps 2-5 are Gibbs draws for a random effects hierarchical model our algorithm inherits the fast convergence properties of these models (see Rosenthal (1995) and Polson (1996)). Since $\underline{\alpha}$ appears in both the demand and supply equations in a nonlinear fashion, convergence is more difficult to assess. Only general results on Metropolis convergence apply (see Roberts and Smith (1993)) and from a practical viewpoint it is sensible to check that the acceptance probability for α_t is at least $10-20\%$. This maintains a proper trade-off between a tighter variance on σ_ξ^2 and σ_η^2 required for accurate estimation and the speed of convergence of the α_t draws.

4.3 Taste Distributions, Parameter Estimates and Elasticities

One of the main advantages of our methodology is that it estimates the distribution of tastes in the market (not just the parameters of a fixed distribution) and also provides parameter and elasticity estimates with little effort. Our model assumes that the distribution of consumer tastes is a mixture of normals of the form:

$$p(\beta) = \int \mathcal{N}(\bar{\beta}, \Sigma_\beta) p(\bar{\beta}, \Sigma_\beta) d\bar{\beta} d\Sigma_\beta$$

The hyperparameters $(\bar{\beta}, \Sigma_\beta)$ are learned from the data via Bayes rule and we obtain a posterior distribution $p(\bar{\beta}, \Sigma_\beta | s, p)$ for them. Our MCMC algorithm provides draws $\{\bar{\beta}^{(g)}, (\Sigma_\beta)^{(g)}\}_{g=1}^G$ from this posterior distribution. These draws allow us to estimate the distribution of tastes given the data by:

$$p(\beta | s, p) = \int \mathcal{N}(\bar{\beta}, \Sigma_\beta) p(\bar{\beta}, \Sigma_\beta | s, p) d\bar{\beta} d\Sigma_\beta$$

which has density estimate:

$$\hat{p}(\bar{\beta} | s, p) \approx \frac{1}{G} \sum_{g=1}^{G} \mathcal{N}\left(\bar{\beta}^{(g)}, (\Sigma_\beta)^{(g)}\right).$$

Since γ has no variance, we compute the marginal posterior distribution $p(\gamma | s, p)$ and use the posterior mean to provide an estimate, namely $\hat{\gamma} \approx \frac{1}{G} \sum_{g=1}^{G} \gamma^{(g)}$.

Finally, we turn to the problem of computing elasticities from the simulation output. One of the key features of our MCMC approach is that the posterior distribution of any nonlinear function like an elasticity can be calculated directly from the simulated draws of our algorithm and Monte Carlo standard errors for these estimates can be computed easily. We illustrate this in the context of elasticities although it applies to any nonlinear function, such as willingness to pay. The own-price elasticity is obtained by integrating over the distribution of consumer tastes (see Nevo (2000)):

$$\frac{\partial \log(s_{jt})}{\partial \log(p_{jt})} = \frac{p_{jt}}{s_{jt}} \int \alpha \frac{s_{jt}(\Theta)\left(s_{jt}(\Theta)-1\right)}{y_t - p_{jt}} p(\Theta) d\Theta \tag{13}$$

where $p(\Theta)$ denotes the probability density function for the parameters. A similar computation for the cross-price elasticity leads to:

$$\frac{\partial \log(s_{jt})}{\partial \log(p_{rt})} = \frac{p_{rt}}{s_{jt}} \int \alpha \frac{s_{rt}(\Theta) s_{jt}(\Theta)}{y_t - p_{rt}} p(\Theta) d\Theta$$

In order to determine the own-price elasticities we use the following MCMC estimator:

$$\frac{p_{jt}}{s_{jt}} \frac{1}{G} \sum_{g=1}^{G} \alpha^{(g)} \frac{s_{jt}(\Theta^{(g)})\left(s_{jt}(\Theta^{(g)})-1\right)}{y_t - p_{jt}}$$

where $(\Theta^{(g)})$ are our MCMC draws. Suppose that: $\int \alpha p_t(\alpha|s_t, p_t) d\alpha < \infty$. Then using the ergodic averaging result of Tierney (1994) it is straightforward to show that this estimator converges to the desired elasticities for almost all starting points of the chain. We now apply our methodology to a simulated data example.

5 Simulated Data and Estimation Results

In this section we examine the performance of our algorithm on a simulated data set. We generate the data based on the underlying assumptions of consumer behavior and firm profit maximization. Specifically, consumers choose the product offering the greatest utility and firms choose prices in a Bertrand-Nash equilibrium. This procedure eliminates the possibility that we are "reverse-engineering" our data to work with our algorithm. The simulation incorporates product-specific effects in both the consumers' utility function and in the marginal cost function. These are observed by consumers and firms but are unobserved characteristics from the econometrician's perspective. This introduces a simultaneity bias if the equations are estimated singly without instruments.

We simulate one-hundred market equilibria based upon consumers facing the choice of two products plus an outside alternative. Products are characterized by three characteristics, x_2, x_3 and the quality measure ξ. There are four cost components, w_1, w_2, w_3 and η. To generate market observations, we draw one-hundred observations for each characteristic for each firm. Table 1 summarizes the characteristics and cost factors for each product.

Table 1: Summary Statistics for Demand and Cost Characteristics

Characteristic	Value	Characteristic	Value
		$w_{1,1}$	-0.945 (0.0902)
$x_{2,1}$	2.6135 (0.3855)	$w_{2,1}$	0.8654 (0.1001)
$x_{3,1}$	2.3212 (0.3922)	$w_{3,1}$	0.6479 (0.0772)
		$w_{1,2}$	-0.9833 (0.0979)
$x_{2,2}$	2.0982 (0.4287)	$w_{2,2}$	0.5162 (0.1138)
$x_{3,2}$	1.4026 (0.3827)	$w_{3,2}$	0.4077 (0.1055)
ξ_1	-0.0206 (0.5453)	η_1	-0.0018 (0.0285)
ξ_2	-0.0537 (0.4963)	η_2	-0.0003 (0.0269)

In each period, consumers maximize

$$u_{ij} = \alpha \log(2 - p_j) + x_{2j}\beta_2 + x_{3j}\beta_3 + \xi_j + \epsilon_{ij}$$

where $\beta_2 \sim \mathcal{N}(1.8, 0.4)$, $\beta_3 \sim \mathcal{N}(2.5, 0.6)$, and $\alpha \sim \mathcal{N}(8, 0.25)$. The price and all characteristics of the outside good are normalized to zero. Our set of consumers is generated with 10,000 draws from the consumer specific, extreme value errors and from the normally distributed taste parameters. We then construct a grid of prices for firm A (the rows) and firm B (the columns) that range from zero to the level of income. For each pair of prices, or cell of the grid, we calculate which of the two goods or the outside good provides the highest utility for each consumer. The shares associated with a set of prices for a given time period are then the proportion of the total number of consumers who choose a product. We construct a grid of prices and shares for each period.

Firms maximize profits in Bertrand-Nash competition, with marginal costs specified by:

$$mc_j = \exp(1w_{1j} + 0.7w_{2j} + 1.2w_{3j} + \eta_j).$$

To find the equilibrium price and shares, we combine the cost information with the share information to calculate the profits of each product for any cell in the grid. We then solve for the Nash equilibrium in prices by finding the cell in the grid in which firm B's choice over columns and firm A's choice over rows coincides (or where this is closest). In order to save computing time, the program solves a coarser grid first and then moves to increasingly finer grids. We stop computing at a grid size of 0.01. This procedure is repeated to find the equilibrium shares and prices for each of the one-hundred time periods. Table 2 reports the mean prices and market shares across time periods of the two products in the simulated data.

Table 2: Summary Statistics for Equilibrium Prices and Shares

	Mean	St.Dev
p_{1t}	1.614	0.0993
p_{2t}	1.165	0.0930
s_{1t}	0.0782	0.1087
s_{2t}	0.489	0.1715

Since the grid used for finding equilibrium prices has a fineness of 0.01, they have an average simulation error equal to the standard deviation of a $U(0, 0.01)$ distribution or 0.00288. This is the added "error," σ_η^2. Since we use an income of 2 and the average price of the products is about 1.3895, log-shares will have an average simulation error of approximately 0.0377 based on a first-order Taylor-series approximation of $\log(s_t)$. This is the added "error," σ_ξ^2.

For estimation, we use the cost characteristics as instruments for the demand equation and the demand characteristics as instruments for the supply relationship. We normalize the product shares by the share of the outside good (i.e. compute the demand intercept) as the estimation routine expects it so that $\beta_1 \sim \mathcal{N}(-6.4646, 0.4269)$. We use the results from a standard instrumented logit regression as starting values for our MCMC algorithm on the demand-side. On the supply-side we use an instrumented regression after substituting the α obtained from the demand-side. For our MCMC estimation, we simulate $5,000$ draws after running our algorithm for a burn-in period of $5,000$ draws. Table 3 presents the estimation results from both methods as well as the true parameter values.

Our MCMC estimates are generally closer to the true parameters and more precisely estimated than the logit estimates, with the exception of two of the cost parameters. The MCMC estimates are within two standard deviations of the true parameters with the exception of α which is within 2.5 standard deviations, γ_1 which is within 2.66, β_1 which is within 2.7, $\sigma_{\beta_1}^2$ which is within 4.1 and $\sigma_{\beta_3}^2$ which is within 2.9. Note that the accuracy of our results depend on how good our instruments are. The instruments are effective to the extent that they are uncorrelated with the unobserved product-specific error we are instrumenting and correlated with the endogenous variables. For the demand-side instruments, z_1 has a correlation of 0.0026 with ξ_j and -0.744 with $\log(y - p)$ and for z_2

Table 3: Estimation Results

Parameter	True Values	Logit Results	MCMC Results
β_1	-6.4646	5.8893 (0.727)	6.7943 (0.122)
β_2	1.800	1.524 (0.680)	1.786 (0.092)
β_3	2.500	2.117 (0.733)	2.662 (0.097)
α	8.000	6.56 (1.08)	7.733 (0.107)
γ_1	1.000	0.978 (0.425)	0.867 (0.050)
γ_2	0.700	0.750 (0.625)	0.684 (0.081)
γ_3	1.200	1.096 (0.893)	1.026 (0.113)
$\sigma^2_{\beta_1}$	0.4269	N/A	0.303 (0.030)
$\sigma^2_{\beta_2}$	0.400	N/A	0.426 (0.056)
$\sigma^2_{\beta_3}$	0.600	N/A	0.418 (0.062)
σ^2_{α}	0.250	N/A	0.200 (0.014)
σ^2_{ξ}	0.00288	N/A	0.000106 (0.0000239)
σ^2_{η}	0.03774	N/A	0.0334 (0.00640)

the correlations are -0.0122 and -0.493. For the supply-side instruments, v_1 has a correlation of -0.069 with η_j and -0.177 with $\log(s)$ and for v_2 the correlations are 0.0597 and -0.458. Since these are far from perfect instruments and there is simulation error in our created data, we do not expect our results to match exactly. Note that the supply-side instruments are not nearly as good as the demand-side instruments which accounts for the lower performance in estimating the supply-side parameters.

Tables 4 and 5 provide further information for comparison. Table 4 presents predicted and actual prices and shares based on period one values of the demand and cost characteristics. As with our parameter estimates, they are generally quite close to the true values. Table 5 displays the true elasticities based upon our simulations as well as the implied own and cross price elasticities from the logit estimation and MCMC estimation. The MCMC estimates are quite close to the true values.

Table 4: Period One Predicted Values (Medians)

	Predicted	True
p_{11}	1.610 (0.0946)	1.694
p_{21}	1.168 (0.0824)	1.264
s_{11}	0.0751 (0.0619)	0.0023
s_{21}	0.228 (.0513)	0.624

Table 5: Period One Elasticities (Medians)

	True Value	Logit Estimate	MCMC Estimate
$\frac{\partial \log(s_{11})}{\partial \log(p_{11})}$	-19.312	-36.183	-19.487
$\frac{\partial \log(s_{21})}{\partial \log(p_{21})}$	-3.944	-4.238	-2.706
$\frac{\partial \log(s_{11})}{\partial \log(p_{21})}$	0.0231	7.027	0.0155
$\frac{\partial \log(s_{21})}{\partial \log(p_{11})}$	1.9501	0.0834	1.307

The model can easily be extended to incorporate demographic data as used in many current applications of random coefficient demand systems. To do this, we would consider a model of the form:

$$u_{ijt} = x'_{jt}\beta + \alpha \log(y_t - p_{jt}) + \xi_{jt} + \epsilon_{ijt}$$

$$\beta = \bar{\beta} + \pi d_{lt} + v$$

where $v \sim \mathcal{N}_K(0, \Sigma_\beta)$ $\xi_{jt} \sim \mathcal{N}\left(\xi_{0t}, \sigma^2_{\xi_t}\right)$ and $\alpha \sim \mathcal{N}(\bar{\alpha}, \sigma^2_\alpha)I_{\alpha>0}$. Here β_i is the marginal utility obtained by sample demographic consumer i from the demand characteristics, $\bar{\beta}$ is the mean marginal utility across all consumers, d_{it} is an $R \times 1$ vector of demographic characteristics for sample demographic consumer i in period t and π is a $K \times R$ matrix of parameters measuring the contribution of each demographic characteristic to marginal utility. As before, the error term v represents the heterogeneity across all consumers in their taste for the characteristics.

Substituting into the utility function and letting q_{jt} denote all interactions between the K demand characteristics and the R demographic characteristics we obtain:

$$u_{ijt} = x'_{jt}\beta + q_{jt}\pi + \alpha \log(y_t - p_{jt}) + \xi_{jt} + \epsilon_{ijt}$$

where π is an $RK \times 1$ vector created by stacking the transpose of the rows of π. Note that the analyst can reduce the dimensionality of q_{jt} and therefore π by eliminating combinations of characteristics and demographics a priori.

Viewing this as a hierarchical model we have:

$$u_{ijt} = x'_{jt}\beta + q_{jt}\pi + \alpha \log(y_{jt} - p_{jt}) + \xi_{jt} + \epsilon_{ijt}$$

$$\pi \sim \mathcal{N}_{RK}(\bar{\pi}, \Sigma_\pi) \ \ , \beta \sim \mathcal{N}(\bar{\beta}, \Sigma_\beta) \ \ , \alpha \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2) I_{\alpha > 0}$$

This model is easy to incorporate into our existing algorithm with the addition of two Gibbs sampling steps to the algorithm: one to draw the parameters π and another to draw a set of data from the demographic distribution. With the straightforward adjustments to the conditional posteriors for β, γ and α to include $q_{jt}\pi$ in the demand-side likelihood, the estimation algorithm proceeds as before and our convergence theorems apply.

6 Conclusion

This paper develops MCMC techniques for estimating likelihood based random coefficient models of market equilibrium. Random coefficient models are increasingly being used to estimate market demand and joint demand-supply equilibrium in situations with differentiated goods where individual level data do not exist. MCMC methodologies provide an attractive framework for these types of problems because of their computational simplicity, finite-sample properties, and flexibility. Parameter distributions are a direct result of the estimation process, and MCMC techniques can easily generate functions of interest based on parameter distributions such as elasticities. The estimation and data analysis are based on likelihood functions and provide finite sample properties. While the researcher provides parametric priors for parameter distributions, the influence of the parametric priors can be controlled by the researcher and the posterior parameter distributions are empirically determined.

Our framework can be expanded in several directions. We describe how to include additional demographic information. Unbalanced panels could also be implemented. Finally, additional information on parameter values from surveys or from other industry sources, can be easily incorporated into the model as priors which reflect the degree of confidence the researcher has in the particular information.

REFERENCES

ANDERSON, S.P., A. DE PALMA, and J.-F. THISSE (1992). *Discrete Choice Theory of Product Differentiation.* Cambridge, MA, MIT Press, Ch. 1-3.

BERRY, S. (1994), Estimating Discrete-Choice Models of Product Differentiation. *Rand Journal of Economics*, 25, 242-262.

BERRY, S., J. LEVINSOHN, and A. PAKES (1995), Automobile Prices in Market Equilibrium, *Econometrica*, 63, 841-890.

BESAG, J. (1974), Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society. Series B*, 36, 192-236.

BESANKO, D., M. PERRY, and R. SPADY (1990), The Logit Model of Monopolistic Competition: Brand Diversity, *Journal of Industrial Economics*, 38, 397-415.

BRESNAHAN, T. (1981), Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 1977-1978. *Journal of Econometrics*, 17, 201-227.

BRESNAHAN, T. (1987), Competition and Collusion in the American Automobile Industry: The 1955 Price War. *Journal of Industrial Economics*, 35, 457-482.

CHAMBERLAIN, G. (1999). Lecture Note 15: Qualitative Choice, Ec2140b, Harvard University Department of Economics, mimeo.

DEGROOT, M.H. (1970). Optimal Statistical Decisions. *New York, NY: McGraw Hill.*

EDWARDS, W., H. LINDMAN and L. J. SAVAGE (1963). Bayesian Statistical Inference for Psychological Research. *Psychological Research*, 70, 193.

GALLANT, R. and TAUCHEN, G. (1996). Which Moments to Match? *Econometric Theory*, 12 657-681.

HASTINGS, W.K. (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57, 97-109.

LIU, J.S., W.H. WONG and A. KONG (1994). Covariance Structure of the Gibbs Sampler with Applications to the Comparison of Estimators and Augmentation Schemes. *Biometrika*, 81, 27-40.

MCFADDEN, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior in P. Zarembka, ed., *Frontiers of Econometrics.* New York, NY, Academic Press.

MENGERSEN, K. and ROBERT, C. (1998). MCMC Convergence Diagnostics: A Review (with discussion) in *Bayesian Statistics 6*, J.M. Bernardo et. al. eds. Oxford University Press, Oxford, England, 399-432.

NEVO, A. (2000). A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics and Management Strategy*, 9, 513-548.

Polson, N.G. (1996). Convergence of Markov Chain Monte Carlo Algorithms (with discussion) in *Bayesian Statistics 5*, J.M. Bernardo et. al. eds. Oxford University Press, Oxford, England, 297-323.

Roberts, G.O. and Smith, A.F.M. (1993). Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society*, 55, 3-23.

Rosenthal, J. (1995). Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 90, 558-566.

Rossi, p., McCulloch, R. and Allenby, G. (1996). The Value of Purchase History Data in Target Marketing. *Marketing Science*, 15, 321-340.

Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions (with discussion). *Annals of Statistics*, 22, 1701-1728.