Bayesian Estimation of Nonlinear Equilibrium Models with Random Coefficients

V. Brian Viard
Cheung Kong GSB, Beijing

Anne Gron
NERA, Chicago

Nicholas G. Polson
Booth School of Business, Chicago [1]

Summary

In this paper we develop a conditional 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.

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 McFadden (1974), Rossi et al (1996), Nevo (2000, 2001), Petrin (2002), Geweke and Keane (2002), Athey and Imbens (2007), Romeo (2007), Burda et al (2008), Jiang et al (2009) and Osborne (2011). 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). 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.

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)).[3]

The demand structure arises from heterogeneous consumers choosing the product that maximizes their utility. Utility is a function of product attributes and disposable income. Goods' characteristics are both observable and unobservable to the econome-

---

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

trician. 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.[4] 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_α^2. ξ_j is a normal random variable with mean zero and variance σ_ξ^2. 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.

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

---

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

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

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

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

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.

Rather than analyze the joint likelihood we analyze a model implicitly defined by the two conditional distributions based upon our demand and supply specifications. We will refer to this as the conditional likelihood system. For technical results on existence of a joint model from a conditionally specified models, see Gelman and Speed (1993), Arnold et al (1999) and Wang and Ip (1998). Zellner et al (2014) provides an alternative direct Monte Carlo approach for Bayesian inference with instrumental variables and also highlights the difficulties involved in calculation of typically highly non-normal posterior distributions that arise in these contexts. Other estimation procedures such as the simulated method of moments (see e.g. Hajivassiliou and McFadden, 1998) can be used in this context.

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.[5] Finally, we discuss the implementation of instrumental variables in our conditional likelihood system.

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 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 $\mathcal{J}(s, p : \Theta, \Lambda)$ be the Jacobian associated with (4) and (6). The joint likelihood function is:

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

---

[5]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.

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 (5). 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 $p(s,p|\Theta)$ by the Hammersley-Clifford Theorem (Besag, 1974). Arnold et al (2001, Theorem 1) provides a recent result on compatible conditional specifying a joint probability model. 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 in Section 3.2.

3.1   Joint Model Specification

Given the equilibrium conditions specified by (4) and (5) we have a conditional likelihood specification given by:

$$\log(s/s_0)|p,\Theta \sim \mathcal{N}\left(X\beta + \alpha\log(y-p) + \xi, \sigma_\xi^2 I_J\right) \tag{7}$$

$$\log\left(p + \alpha^{-1}\Lambda(s)(y-p)\right)|s,\Lambda,\Theta \sim \mathcal{N}\left(w\gamma, \sigma_\eta^2 I_J\right) \tag{8}$$

where we have stacked the variables here as:

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

The demand-side conditional likelihood is then proportional to:

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

where $g(p,y,\alpha,\beta) = \log(s_0) + X\beta + \alpha\log(y-p) + \xi$.

The supply-side conditional likelihood is proportional to:

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

where $h(p,s,y,\alpha) = \log(p + \alpha^{-1}\Lambda(s)(y-p))$ and $||x||^2 = x'x$ and T is the number of markets served.

8

Due to the invertibility of $h(p, s, y, \alpha) = \log(p + \alpha^{-1}\Lambda(s)(y - p))$ in p given s and y, we can write the model specification directly in terms of the conditional distributions of shares and prices: denoted by $(s|p, \Theta)$ and $(p|s, \Theta, \Lambda)$. We have conditional distributions where $Z_1 \sim \mathcal{N}(0, I_J)$ and $Z_2 \sim \mathcal{N}(0, I_J)$ given by

$$(s|p, \Theta) \sim e^{g(p, y, \alpha, \beta) + \sigma_\xi Z_1} \tag{9}$$

$$(p|s, \Theta, \Lambda) \sim (1 - \alpha^{-1}\Lambda(s))^{-1}\left(e^{w\gamma + \sigma_\eta Z_2} - \alpha^{-1}\Lambda(s)y\right) \tag{10}$$

The distribution of shares given prices is log-normal and that of prices given shares a shifted log-normal. Therefore, given parameters, Θ, and the competition matrix, Λ, we know the distribution of shares given prices and vice-versa the distribution of prices given shares. We now show that this is sufficient to define a joint probability distribution for shares and prices given parameters and form of competition.

The Clifford-Hammersley-Besag formula allows us to write, suppressing the conditioning on Θ and Λ, for any pair $(s^{(0)}, p^{(0)})$ the joint density $p(s, p)$ as a product of conditional densities

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

There is a mild positivity condition which requires that for each point in the sample space, $p(s, p)$ and the marginals $p(s)$ and $p(p)$ have positive mass. Given our log-normal specification this holds for our conditional specification. Hence we have a well defined joint likelihood, $\mathcal{L}(s, p|\Theta, \Lambda)$, which can be computed as:

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

The prior $p(\Theta, \Lambda)$ is well-defined and, by construction, so is the posterior. We simply combine the likelihood $\mathcal{L}(s, p|\Theta, \Lambda)$ with the random effects distribution of the parameters, $p(\Theta, \Lambda)$. Bayes' rule to yield the posterior distribution of the parameters given the data:

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

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

Our problem, however, is one of inverse probability: given shares and prices in a market equilibrium can calculate the posterior distribution the parameters and the type of competition (e.g. Bertrand or Cournot) given shares and prices? MCMC methods

provide a natural approach to solving this system. These methods break the problem up into a number of simpler problems. 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. Section 4 provides the details.

3.2 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 $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$. [6]

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:

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

---

[6]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.

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 3.4, 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 $\sigma_\xi^2 z z'$ and $\sigma_\eta^2 v v'$ 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.

3.3 Application to Random Coefficient Demand and Oligopoly Supply

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:

$$\log(s_t) = \alpha \log(y_t - p_t) + X_t \bar{\beta} + e_t^D \tag{11}$$

where

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

$e_t^D \sim \mathcal{N}(0, X_t \Sigma_\beta X_t' + \sigma_\xi^2 I)$ 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 $x_{0t} \Sigma_\beta x_{0t}'$.[7] 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.

---

[7] 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}.

Similarly, the observed vector of prices of products in time period t, p_t, is a vector of random variables implicitly defined by:[8]

$$\log[p_t + \alpha^{-1}\Lambda_t(s_t)(y_t - p_t)] = w_t\gamma + e_t^S \tag{12}$$

where

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

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 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.

3.4 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 (11) and (12) by multiplying the demand side through by $z_t = [z_{1t} \dots z_{Jt}]$ and the supply-side by $v_t = [v_{1t} \dots v_{Jt}]$ to get:

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

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

---

[8]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 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:

$$z_t \log(s_t) | v_t \log(A_t), \alpha, \bar{\beta}, \Sigma_\beta, y_t \sim \mathcal{N}(\alpha z_t \log(y_t - p_t) + z_t X_t \bar{\beta}, \Sigma_D) \qquad (13)$$

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

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:

$$\alpha \sim \mathcal{N}(a_0, A_0^{-1})I_{\alpha>0} \quad \sigma_\alpha^2 \sim \mathcal{W}(a_2, A_2^{-1})$$

$$\bar{\beta} \sim \mathcal{N}(b_0, B_0^{-1}) \quad \Sigma_\beta \sim \mathcal{W}(b_2, B_2^{-1})$$

$$\gamma \sim \mathcal{N}(g_0, G_0^{-1})$$

where $\mathcal{W}$ denotes the inverted Wishart distribution.[9] 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:

$$z_t \log(s_t)|p_t, \Theta \sim \mathcal{N}(\alpha z_t \log(y_t - p_t) + z_t x_t \beta, z_t z_t' \sigma_\xi^2)$$

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

$$\bar{\beta} \sim \mathcal{N}(b_0, B_0^{-1}) \quad \Sigma_\beta \sim \mathcal{W}(b_2, B_2^{-1}) \quad \sigma_\alpha^2 \sim \mathcal{W}(a_2, A_2^{-1}).$$

---

[9] 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.

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:[10]

$$\mathcal{L}(s, p|\Theta) = \prod_{t=1}^{T} p(\log(s_t)|p_t, \Theta)p(p_t|\Theta). \tag{14}$$

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 (14) 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:

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

---

[10]We make a simplifying assumption here: $p^*(\Theta) = p(p_t|\Theta)/p(p_t|log(s_t), \Theta) \propto 1$. 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: $p(log(s_t), p_t|\Theta) = p(log(s_t)|p_t)p(p_t|\Theta)$.

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:

$$A_3 = A_S + A_D + \sigma_\alpha^2$$

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

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 $\bar{\alpha}|\underline{\alpha}, \sigma_\alpha^2, s, p$ and $\sigma_\alpha^2|\underline{\alpha}, \bar{\alpha}, s, p$. Standard conjugate theory (see DeGroot, 1970) implies:

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

where

$$A_1 = T\sigma_\alpha^2 + A_0$$

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

and

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

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 $\Sigma_\beta | \bar{\beta}, \beta, \underline{\alpha}, s, p$. 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

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

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

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 $\bar{\beta}$.

17

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

with
$$R_\gamma = \sigma_\eta^2 \text{diag}\, (v_1 v'_1, \ldots, v_T v'_T)$$

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 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)):

$$\frac{\partial \log(s_{jt})}{\partial \log(p_{jt})} = \frac{p_{jt}}{s_{jt}} \int \alpha \frac{s_{jt}(\Theta)(s_{jt}(\Theta) - 1)}{y_t - p_{jt}} p(\Theta) d\Theta \tag{15}$$

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   ESTIMATION RESULTS

5.1. Simulated Data

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

Table 3: Estimation Results

Parameter	True Values	Logit Results	MCMC Results
β_1	-6.4646	5.8893	6.7943
		(0.727)	(0.122)
β_2	1.800	1.524	1.786
		(0.680)	(0.092)
β_3	2.500	2.117	2.662
		(0.733)	(0.097)
α	8.000	6.56	7.733
		(1.08)	(0.107)
γ_1	1.000	0.978	0.867
		(0.425)	(0.050)
γ_2	0.700	0.750	0.684
		(0.625)	(0.081)
γ_3	1.200	1.096	1.026
		(0.893)	(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)

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, $\sigma^2_{\beta_1}$ which is within 4.1 and $\sigma^2_{\beta_3}$ 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 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:

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

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

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.

5.2 U.S. Automobile Market Equilibrium Data

We use our methodology to estimate demand and supply equilibrium in a balanced panel of 58 automobile models over the five year period, 1988-1992.[11] This sample

---

[11] Car manufacturers often sell multiple versions of each car model. Our analysis considers the most basic four-door version of each car. If the particular model is not sold in a four door version, such as some sports cars, we instead look at the two door version. We confine our sample to that set of nameplates for which there are observations in all the years.

represents between 77% and 64% of the total market quantity.[12] We assume that consumers and producers consider different prices in their respective decisions. Since a car is a durable, capital good, consumers consider the annual flow cost when making the purchase decision. Producers, on the other hand, are producing a capital good and are setting the price of the stock. We also allow the value of the outside alternative to differ over time. In our application it is obvious why we do this; furthermore, the value of the alternative good likely differs across time and markets in other applications as well. Even if we included all car models and light trucks produced over the period, the increased use of fleet sales and leasing arrangements has altered the market for used cars, an important alternative good. For other consumer products, one can readily imagine that consumers in different regional markets may place different value on the outside good.[13] Finally, we allow the marginal utility of income to be a random variable as well.

Our estimation data includes automobile characteristics, list prices, quantities, factor costs and car loan information to transform the list price into an annual flow cost. List prices, quantities and characteristics such as air-conditioning, horsepower, weight, size, and miles per gallon come from Automotive News: Market Data Book.[14] For our demand estimation, we follow existing specifications and include horsepower divided by weight $\mathsf{HP/WT}$, miles per dollar MPD, air conditioning Air and size (length×width) Size. To allow for a changing value of the outside good, we include time indicator variables D_t in our demand side estimation. This results in the following demand-side specification (before instrumenting):

$$\ln s_{jt} = \alpha \ln(y_t - p_{jt}) + \sum_{t=1}^{T} \beta_{0t}\mathsf{D}_t + \beta_1\mathsf{Air}_{\mathsf{jt}} + \beta_2\mathsf{MPD}_{\mathsf{jt}} + \beta_3\mathsf{Size}_{\mathsf{jt}} + \beta_4\mathsf{HP/WT}_{\mathsf{jt}} + \psi_{jt} + e_{jt}^S$$

Our pricing equation differs somewhat from other specifications in the recent automobile literature in that we combine automobile characteristic and factor price information in our marginal cost specification. Following the past literature, we include Air, $\log(\mathsf{HP/WT})$ and $\log(\mathsf{Weight})$. We also include a time trend Trend to pick up technical change during the period. In addition to model characteristics, we include a cost factor, Cost, which represents a value-weighted index of the labor costs associated with a particular model. In particular, we specify:

$$\mathsf{Cost}_{jt} = \%\mathsf{USContent}_{jt}\mathsf{relw}_{US,t} + (1 - \%\mathsf{USContent}_{jt})\,\mathsf{relw}_{F,t}$$

where F indexes the country of origin for non-US based manufacturers, % $\mathsf{US\ Content}$ is the US value-added as reported in the annual reports of the Foreign Trade Zones

---

[12]We use the balanced panel for programming purposes, the estimation programming is more direct with the balanced panel but can be extended to include unbalanced panels as well. Note also that we adjust our estimation to account for the fact the value of the outside good changes over time, which would be the case even if we had an unbalanced panel.

[13]For example, consider breakfast cereals with the alternatives being toast, muffins or big breakfasts. You can imagine regional tastes and customs may make some of these more attractive, or alternatively, the value of cereal versus other breakfast choices may be different in different areas.

[14]We thank Jim Levinsohn for generously sharing these data with us.

Board [15] and relw is an index of relative manufacturing wage rates for the US or for the manufacturer's country of origin taken from the Bureau of Labor Statistics Monthly Bulletin. For a given set of characteristics, a change in factor prices will increase or decrease marginal cost.[16] Therefore our marginal cost specification is:

$$mc_{jt} = \mathsf{Cost}_{\mathsf{jt}}{}^{\gamma_1} \mathsf{Trend}_{\mathsf{t}}{}^{\gamma_2} \mathsf{HP/WT}_{\mathsf{jt}}{}^{\gamma_3} \mathsf{Weight}_{\mathsf{jt}}{}^{\gamma_4} \exp\left(\gamma_0 + \gamma_5 \mathsf{Air}_{\mathsf{jt}} + \eta_{jt}\right).$$

Our price measures are based on list prices rather than transaction prices, like other recent work in this area.[17] Our approach differs from these others in that we specify both a consumer price, P_C and a producer price P_P. In the case of a capital good such as a car, a consumer will base the purchase decision on the flow cost of "consuming" the services from the car. This price measure is also a more appropriate comparison for our income measure which is the annual income flow. In terms of the purchase price, we can approximate that flow cost, denoted P_C, by calculating the annual loan payments a buyer would have to make given current automobile loan interest rates, r, loan duration, n, and list price, P_P. We construct this flow cost with loan rate data from the Federal Reserve Board of Governors as: $P_C = \left(1 - (1+r)^{-n}\right)^{-1} r P_P$ where r is the relevant loan rate. An automobile manufacturer, on the other hand, is producing a capital good and sets a price for the automobile itself. Therefore we use the list price for the producer price. This difference between producer and consumer prices results in an additional adjustment to the pricing equation. All price and cost measures are deflated to 1988 real prices using the consumer price index deflater.

For the full system we also take into consideration a dealer mark-up. The manufacturer sells the automobile to the dealer at a wholesale price, P_P, who then sells it to the consumer at a retail price, P_R. If d_j denotes the dealer mark-up for good j, taking account of the dealer mark-up we will estimate:

$$\ln\left((I_J - \alpha^{-1}\Lambda_t(s_t))P_{R,t} + \alpha^{-1}k^{-1}(1+d_j)^{-1}\Lambda_t(s_t)Y_t\right) = W_t\gamma + \eta_t + e_t^P$$

where $k^{-1} = \left(1 - (1+r)^{-n}\right)r^{-1}$.

We also need to specify instruments for prices and shares, as noted in the previous sections. As in other specifications, we exploit the panel nature of the data and our assumptions that the observed characteristics are exogenous. Therefore, functions of characteristics of other products are instruments for price and share. For model j we use

---

[15]US and foreign automobile makers who assemble automobiles in the US do so in a designated Foreign Trade Zone (FTZ). Operations within a FTZ receive more favorable tariff treatment on imported intermediate inputs. For additional information on FTZ's and this cost measure, see Gron and Swenson (2000).

[16]We focus on labor costs because labor is the primary input which is location specific.

[17]While transactions level prices would be preferred, they are typically not available when individual transactions are not observed. Our list price variable is the same price variable as that used in studies conducted by Berry, Levinsohn and Pakes (1995), Fershtman and Gandal (1998), Petrin (2002) and Verboven (1996).

the characteristics themselves and the counts and sums of the characteristics of two groups of products as our instruments: other products produced by the same manufacturer and products produced by all other manufacturers.

The dimensionality of our variance-covariance matrix equals the number of instruments and with only five time periods requires that we impose some structure on the demand and supply covariance matrices. We impose structure by using a shrinkage prior distribution that shrinks to a heteroskedastic matrix $\sigma_S^2 Z'Z$ for demand and $\sigma_P^2 V'V$ for supply.[18]

In our estimation we ran three specifications, in increasing order of generality, to determine the influence of the model choice on the results. Table 1 presents estimated demand parameters from these three models. Model I, the Logit Model, corresponds to a discrete choice demand system with unobserved product characteristics but without random coefficients. It is estimated using GLS. Since it is based on a logit demand model, the usual consequences of the IIA assumption follow. The inclusion of unobserved product characteristics, however, somewhat reduces the dependency on the logit error term. Model II, the Demand-Only Model, introduces two differences relative to the logit model. First, the demand coefficients are random allowing for more sensible own- and cross- price elasticities. Second, it imposes the restriction $E(Z'\psi) = 0$ through our prior specification[19]. Model III, the Full-System Model, is our random coefficient model with both demand and supply equations jointly estimated and $E(Z'\psi) = E(V'\eta) = 0$. All models include time dummies to allow the value of the outside good to vary over time.

As Table 6 shows, the parameter estimates are relatively unchanged in moving from Model I to II. For the random coefficient demand and supply models we report posterior means and standard deviations as the underlying posteriors for the parameters are nearly normal.[20] However, the elasticities generated by the two models differ dramatically. The own price elasticities from the Demand-Only model are larger in magnitude than those from the Logit model (see Table 7). The cross-price elasticities from the Demand-Only model are not the same for all models with respect to a given model's price, a well-known flaw of the logit (see Table 8). The cross-price elasticities from the Demand-Only model vary somewhat more reasonably with car pairings.

---

[18]In our estimation of the U.S. auto market we set $\sigma_S^2 = .05$ and $\sigma_P^2 = .07$ which are small relative to the mean log-share of -6.27 and mean price of \$14,231. Prior distributions on all other parameters were set to be highly diffuse.

[19]This is accomplished by setting the appropriate elements of b_0 to zero.

[20]Our simulation period consisted of $5,000$ iterations after a burn-in period of $5,000$ iterations. As is typical with MCMC procedures, a number of parameters in the algorithm can influence the mixing of the Markov Chain. For example, the variance of our candidate distribution will affect the acceptance probability of the chain. Too low a probability will lead to poor convergence. We tried different starting values to check the sensitivity of the algorithm to initial conditions. Consumer prices were used throughout. The characteristics variables MPD, Size, HP/WT and Weight are normalized by their standard deviation. The demand equation includes indicator variables by year and the supply equation includes a time trend. Logit regression uses the average income.

Table 6: Demand Side Posterior Estimates: Models I, II, III

Model	I	Logit	II	Demand-Only	III	Full-System
Demand	Mean	S.E.	Mean	S.E.	Mean	S.E.
Air	0.353	0.338	0.259	0.328	1.114	0.304
MPD	-0.156	0.163	-0.148	0.160	-0.454	0.154
Size	0.743	0.110	0.810	0.101	0.453	0.087
HP/WT	-0.136	0.110	-0.108	0.100	-0.183	0.100
ln(y-p)	3.247	1.855	3.660	1.458	13.409	0.270
Const	-43.205	17.818	-47.986	13.960	- 140.371	3.066

Table 7: Own-Price Elasticities

Model	I	II	III
	Logit	Demand Only	Full System
Ford Tempo	-0.3521	-0.9872	-5.0806
Ford Taurus	-0.4763	-0.7313	-3.1579
Toyota Camry	-0.4319	-0.2761	-1.0539
Toyota Celica	-0.4650	-0.8719	-1.6662
Nissan Maxima	-0.7257	-1.4060	-11.7942
Acura Legend	-0.9756	-0.9020	-3.8154
Mazda 323	-0.2351	-1.6361	-7.0011
Mercedes-Benz 300D	-2.2225	-2.1340	-2.7384
BMW 733/35/40i	-3.2696	-7.5206	-10.5790

Comparing Models II and III illustrates the benefits of combining demand and supply information in a likelihood based analysis. As shown in Table 6, the marginal utility of income, α, estimated by the Full-System model is greater than that from the Demand-Only model. In addition, the coefficient on Air moves from being insignificant to being significant and of the expected sign, while Size remains significant and of the expected sign. The coefficient on HP/WT remains of the unexpected sign but only marginally significant. Table 9 provides parameter estimates for the supply equation under the Full-System model. Air is significant and of the expected sign. Cost, log(HP/WT) and log(Weight) are of the expected sign but are not significant.

The effects on elasticities in moving from Model II to Model III are dramatic. Table 7 reveals a vast improvement in own-price elasticity estimation from including the supply-side information. Many own-price elasticities move from being inelastic to elastic

Table 8: Cross Price Elasticities for Model II Demand-Only

Model	Ford Tempo	Ford Taurus	Toyota Camry	Toyota Celica	Nissan Maxima	Acura Legend	Mazda 323	MB 300	BMW 733/5i
Ford Tempo	-0.9872	0.0304	0.0095	0.0254	0.0556	0.0244	0.0625	0.0651	0.2522
Ford Taurus	0.0327	-0.7313	0.0071	0.0213	0.0371	0.0211	0.0396	0.0453	0.1838
Toyota Camry	0.0102	0.0071	-0.2761	0.0069	0.0129	0.0068	0.0146	0.0158	0.0595
Toyota Celica	0.0071	0.0055	0.0018	-0.8719	0.0115	0.0061	0.0122	0.0136	0.0507
Nissan Maxima	0.0127	0.0079	0.0027	0.0095	-1.4060	0.0083	0.0213	0.0205	0.0694
Acura Legend	0.0022	0.0018	0.0006	0.0020	0.0034	-0.9020	0.0035	0.0045	0.0172
Mazda 323	0.0100	0.0059	0.0022	0.0070	0.0149	0.0062	-1.6361	0.0154	0.0561
Mercedes-Benz 300	0.0012	0.0008	0.0003	0.0009	0.0017	0.0009	0.0018	-2.1340	0.0091
BMW 733/35/40i	0.0014	0.0010	0.0003	0.0010	0.0017	0.0011	0.0020	0.0027	-7.5206

Table 9: Supply Side Posterior Estimates: Model III Full-System

Model	III	Full-System
Supply	Mean	S.E.
Air	0.819	0.336
ln(HP/WT)	0.210	0.147
ln(Cost)	0.079	0.120
ln(Weight)	0.074	0.096
Const	5.856	5.450
Trend	-0.073	0.039

as we would expect under an assumption of profit-maximizing firms.

The logit elasticities are less elastic than those from the random coefficient full system. Using the full system estimates, the magnitude of the own-price elasticity of demand tends to be increasing in the price of the car. Goldberg (1995) reports similar findings in estimation using transaction level data and buyer characteristics.[21] One explanation is that lower priced cars are closer to "necessities" and therefore have lower elasticity of substitution while higher priced cars behave as luxuries. While we expect a $1000 change in the price to have less impact on demand for a luxury car, we would not necessarily expect a higher priced car to have a lower own-price elasticity of demand. Another explanation is that our result is mainly due to assumptions about the income

---

[21]Goldberg (1995) also reports average own-price elasticities by automobile category. These average elasticities are increasing in price for car classes with the exception of luxury and sports and speciality models. These classes are dealt with somewhat differently in her estimation.

distributions. Own-price elasticity is defined as $\alpha(s-1)p/(y-p)$. For high priced cars $s-1 \approx -1$ and the different estimates of α affect the level and not the relative position of elasticities. Hence the dominant term is $(y/p-1)^{-1}$ which clearly depends on the income distributions and not the particular random utility specification.

Table 10 shows a sample of cross-price elasticities for pairings between 9 car models. While there are exceptions, cross-price elasticities tend to be greater between pairings of cars which would be expected to act as good substitutes and lower for those which are more distant substitutes. Cross-price elasticities can be greatly affected by the random coefficient specification used. This is due to the fact that a cross-price elasticity is given by $\alpha sp/(y-p)$ where the shares s will depend on the estimated distribution of the characteristics β's and unobservables ψ's. The numerator is therefore highly dependent on the estimates of the β's and ψ's and can change dramatically across specifications. Also the distribution of α can change dramatically shifting the elasticities. To compute elasticities we take into account the conditioning on the set $\{y_t > p_t\}$.[22] We adjust for this and compute the marginal elasticities using

$$E\left(\frac{\partial \ln s_{jt}}{\partial \ln p_{jt}}\right) = E\left(\frac{\partial \ln s_{jt}}{\partial \ln p_{jt}}|y_t > p_t\right) P(y_t > p_t)$$

Table 10: Cross Price Elasticities for Model III Full-System

Model	Ford Tempo	Ford Taurus	Toyota Camry	Toyota Celica	Nissan Maxima	Acura Legend	Mazda 323	MB 300	BMW 733/5i
Ford Tempo	-5.0806	0.2106	0.0526	0.0769	0.6672	0.1448	0.3594	0.1069	0.4163
Ford Taurus	0.2257	-3.1579	0.0331	0.0437	0.3486	0.1016	0.2301	0.0783	0.2050
Toyota Camry	0.0527	0.0308	-1.0539	0.0121	0.1212	0.0309	0.0574	0.0206	0.0584
Toyota Celica	0.0219	0.0118	0.0033	-1.6662	0.0451	0.0115	0.0262	0.0134	0.0220
Nissan Maxima	0.1287	0.0621	0.0217	0.0325	-11.7942	0.0666	0.1265	0.0388	0.1224
Acura Legend	0.0116	0.0075	0.0022	0.0034	0.0275	-3.8154	0.0149	0.0051	0.0162
Mazda 323	0.0569	0.0322	0.0092	0.0147	0.1063	0.0295	-7.0011	0.0207	0.1059
Mercedes-Benz 300D	0.0019	0.0010	0.0003	0.0005	0.0034	0.0010	0.0020	-2.7384	0.0032
BMW 733/35/40i	0.0010	0.0006	0.0001	0.0002	0.0023	0.0006	0.0012	0.0003	-10.5790

An advantage of our approach is the ability to obtain distributions for the random coefficients in the model. Our estimation method is not based on moment matching as it is likelihood based. Therefore, particularly for the full system, we should not expect the model to match exactly the observed market shares and prices. However, the fit is rather close in the full system. The mean-squared error of the logarithm of predicted

---

[22]The distribution of income y_t is taken from a log-normal distribution with mean \$24K and small standard deviation. We directly simulate the probabilities $P(y_t > p_t)$ conditional on purchase.

shares is .000204 compared to an average actual log share of .0036 and for log predicted prices is .105 relative to 9.42. Our likelihood based approach optimally trades-off the information in the supply and demand systems when estimating the underlying parameter distributions. We see that the full system matches the observed prices more closely than it matches the observed market shares. This is one of the main advantages of a likelihood based method over a moment matching approach.

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.

ARNOLD, B.C., E. CASTILLO and J.M. SARABIA (1999). *Conditional Specification of Statistical Models.* Springer-Verlag.

ARNOLD, B.C., E. CASTILLO and J.M. SARABIA (2001). Conditionally Specified Distributions: An Introduction. *Statistical Science*, 16(3), 249-274.

ATHEY, S. and G.W. IMBENS (2007). Discrete Choice models with multiple unobserved choice characteristics. *International Economic Review*, 48(4), 1159-1192.

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.

BURDA, M., M. HARDING and J. HAUSMAN (2008). A Bayesian mixed logit-probit model for multinomial choice. *Journal of Econometrics*, 147, 232-246.

CHAMBERLAIN, G. (1999). Lecture Note 15: Qualitative Choice. 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.

FERSHTMAN, C. and N. GANDAL (1998). The Effect of the Arab Boycott on Israel: The Automobile Market, *Rand Journal of Economics*, 29, 193-214.

GELMAN, A. and T.P. SPEED (1993). Characterising a joint probability distribution by conditionals. *Journal of Royal Statistical Society, B*, 55, 185-188.

GEWEKE, J. and M. KEANE (2002). Bayesian Inference for Dynamic Discrete Choice Models without the need for Dynamic Programming. *Simulation-Based Inference in Econometrics*, CUP, 100-131.

GOLDBERG, P. (1995), Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry, *Econometrica 63(4): 891-951.*

GRON, A. and D. SWENSON (1998). Cost Pass-Through in the US Automobile Market. *Review of Economics and Statistics*, 82(2), 316-324.

HAJIVASSILIOU, V. and D. MCFADDEN (1998). The Method of Simulated Scores for Estimation of LDV Models. *Econometrica*, 66(4), 863-896.

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

JIANG, R., P. MANCHANDA and P.E. ROSSI (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics*, 149, 136-158.

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.

Nevo, A. (2001). Measuring Market Power in the Ready-to-eat Cereal Industry. *Econometrica*, 69(2), 307-342.

Osborne, M. (2011). Consumer Learning, Switching Costs and Heterogeneity: a Stuctural Examination. *Marketing and Economics*, 9(1), 25-70.

Petrin, A. (2002). Quantifying the benefits of new products: the case of the Minivan. *Journal of Political Economy*, 110(4), 705-729.

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.

Romeo, C. (2007). A Gibbs sampler of Mixed Logit Analysis of Differentiated Product markets using Aggregate Data. *Computational Economics*, 29(1), 33-68.

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.

Trajtenberg, M. (1989). The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners. *Journal of Political Economy, 94, 444-479*.

Verboven, F. (1996). International Price Discrimination in the European Car Market. *Rand Journal of Economics, 27, 240-268*.

Wang, Y.J. and E.H. Ip (2008). Conditionally specified continuous distributions. *Biometrika*, 95(3), 735-746.

Zellner, A., T. Ando, N. Basturk, L. Hoogerheide and H. van Dijk (2014). Instrumental Variables, Errors in Variables and Simultaneous Equations Models: Applicability and Limitations of Direct Monte Carlo. *Econometric Reviews*, 33, 3-35.