Quantitative Marketing and Economics, 1, 251–275, 2003.

Bayesian Analysis of Simultaneous Demand and Supply

SHA YANG
Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012
E-mail: shayang@stern.nyu.edu

YUXIN CHEN
Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012
E-mail: ychen@stern.nyu.edu

GREG M. ALLENBY*
Fisher College of Business, Ohio State University, 2100 Neil Avenue, Columbus, OH 43210–1144
E-mail: allenby.1@osu.edu

Abstract. In models of demand and supply, consumer price sensitivity affects both the sales of a good through price, and the price that is set by producers and retailers. The relationship between the dependent variables (e.g., demand and price) and the common parameters (e.g., price sensitivity) is typically non-linear, especially when heterogeneity is present. In this paper, we develop a Bayesian method to address the computational challenge of estimating simultaneous demand and supply models that can be applied to both the analysis of household panel data and aggregated demand data. The method is developed within the context of a heterogeneous discrete choice model coupled with pricing equations derived from either specific competitive structures, or linear equations of the kind used in instrumental variable estimation, and applied to a scanner panel dataset of light beer purchases. Our analysis indicates that incorporating heterogeneity into the demand model all but eliminates the bias in the price parameter due to the endogeneity of price. The analysis also supports the use of a full information analysis.

Key words. discrete choice model, endogeneity, heterogeneity, hierarchical Bayesian analysis

JEL Classification: M31, C11, C33, C35

1. Introduction

A growing body of literature has addressed the importance of accounting for the simultaneity of demand and supply in studies of consumer behavior. The rationale is that firms often set prices and allocate resources based on their expectation of the market's response. When this occurs, variables that are typically assumed as explanatory in the demand model (e.g., the 4 p's) are determined from within the system of study, and are dependent on some of the same parameters present in the demand relationship (see Manchanda et al., 2003). Sellers, for example, often set

*Corresponding author.

price after taking into account consumer preferences, price sensitivity and variables (e.g., promotional campaigns) that may not be observed by the researcher, which result in an endogenous relationship between demand and supply that is not well represented by models that condition on price. Failure to account for price endogeneity leads to estimation bias as shown in both aggregate and disaggregate data (Berry, 1994; Villas-Boas and Winer, 1999).

In this paper, we develop a Bayesian method of estimating a heterogeneous demand and supply model with household panel data. The proposed method is general in a sense that it can be used to estimate parameters in a wide variety of simultaneous equation systems. The method is developed within the context of a heterogeneous discrete choice model coupled with pricing equations derived either from alternative competitive structures or an instrumental variable specification similar to that used in limited-information analyzes. As with all Bayesian methods, the proposed method facilitates exact, finite-sample inference and does not rely on asymptotic results.

Current likelihood-based approaches to jointly estimating simultaneous demand and supply systems have trouble incorporating unobserved heterogeneity into the model specification, particularly when the supply-side (e.g., pricing) model is derived from specific competitive assumptions. Researchers have instead focused on the common shocks, and have only partially controlled for consumer heterogeneity. Past purchase histories and demographics variables have been incorporated into the model specification, but heterogeneous consumer response variables such as prices have not (Villas-Boas and Zhao, 2003; Draganska and Jain, 2002).

Alternative approaches have been proposed to deal with the computational complexity of conducting the analysis of simultaneous demand and supply. One approach involves writing the supply equation as a linear model with an error term that is correlated with the demand shocks (Villas-Boas and Winer, 1999). This approach results in significant simplification of the likelihood function, but does not facilitate study of the nature of the supply side relationship. A second approach, due to Berry (1994), proceeds by first estimating a set of brand- and time-specific fixed-effects that agree with observed market shares. These fixed-effect estimates are then used as dependent variables in a subsequent regression analysis to estimate brand characteristics and consumer price sensitivity, using instruments to control for endogeneity. Berry et al. (1995) use this method to investigate automobile pricing, Sudhir (2001a) uses it to study the automobile market competition, and Goolsbee and Petrin (2001) use it to investigate the introduction of direct broadcast satellite as an alternative to cable television. Variants of this approach include the Chintagunta et al. (2002) who use MLE approach to estimate the set of brand- and time-specific fixed-effects, the Petrin and Train (2002) who propose including residuals, obtained from regressing price on a set of instruments, in the subsequent regression analysis, and Kuksov and Villas Boas (2001) who use predicted value of the endogenous variable from instruments followed by method-of moments estimator.

While these alternative approaches facilitate the inclusion of heterogeneity and a richer supply-side model specification, the efficiency and small sample properties of these estimators are unknown. Small-sample properties of estimators are of interest to researchers in marketing because marketing data is characterized by many "units" of analysis (e.g., households) and few observations per unit. The output from the first step carries uncertainty that is difficult to track in the second step, and this uncertainty is likely to be large given the small number of observations available per unit. Moreover, it is not clear how large a sample is required for asymptotic properties of the estimator to apply, and it can be difficult to identify suitable instrumental variables for analysis.

We illustrate our method using a household scanner panel data of light beer purchases. The analysis indicates that incorporating parameter heterogeneity into the demand model all but eliminates the reported bias in the price parameter due to the endogeneity of price. The analysis also supports the use of pricing equations derived from a specific competitive structure, with the evidence about parameter bias sensitive to the presence of both heterogeneity and the supply side specification. The results underscore the importance of the model specification. The remainder of the paper is organized as follows: in the next section we specify the likelihood and discuss the estimation procedure. An empirical application is then provided in Section 3, and Section 4 offers a discussion and concluding comments.

2. Simultaneous demand and supply

The estimation of simultaneous demand and supply with household heterogeneity is challenging. In a full information analysis, where specific assumptions are made about the nature of competition among firms, the supply-side equation is a complex function of household-level parameters and common error terms, or shocks. To date, likelihood-based approaches have not been developed for models with unobserved heterogeneity specified as random-effects. The reason is that a frequentist (i.e., non-Bayesian) approach to analysis does not view the shocks and random-effects as parameters, and analysis proceeds by first integrating them out of the likelihood function. This integration is computationally demanding when the shocks and random-effects are jointly present, and researchers have instead controlled for consumer heterogeneity by incorporating past purchases, or other exogenous (e.g., demographic) variables, into the model specification (Villas-Boas and Zhao, 2003; Draganska and Jain, 2002). In the context of demand analysis using aggregate level data, Chintagunta (2001) has shown that it is important to control for both price endogeneity and heterogeneity to avoid potential biases in demand side parameter estimates.

Our Bayesian approach to estimating simultaneous demand and supply does not require the immediate integration of the demand shocks and random-effects. In a Bayesian analysis, these unobservable quantities are treated as latent variables and treated as objects of inference. To a Bayesian, there is no difference between model parameters and other unobserved latent variables. When Bayesian analysis is coupled with Markov chain Monte Carlo estimators, in which draws are generated from the full conditional distribution of all model parameters, computational simplification arises from the ability to condition on the latent shocks and random-effects.

In this section, we first set notation by developing the demand and supply models used in our study of consumer brand choice and retailer pricing. Consumers are assumed to make brand choice decisions according to a standard discrete choice model. On the supply side, we develop specifications derived from profit-maximizing assumptions made about manufacturer and retailer behavior, and also investigate a linear model specification similar to that employed when using instrumental variables. We then turn to the primary purpose of our paper—the Bayesian analysis of simultaneous demand and supply.

2.1. *Demand model*

The disaggregate demand function is specified as a logistic normal regression model. This model has been used in the analysis of demand data by Allenby and Lenk (1994), and in the analysis of simultaneous demand and supply by Berry et al. (1995), Villas-Boas and Winer (1999), and Chintagunta et al. (2002). Suppose we observe purchase incidences and choices (y) for a group of individuals ($i = 1, \ldots, I$) for J brands ($j = 0, \ldots, J$) in a product category over T time periods ($t = 1, \ldots, T$). The utility of consumer i for brand j at time t is specified as:

$$u_{ijt} = \beta_i' x_{jt} + \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt}, \tag{1}$$

and in the case of no purchase from the J available brands, we denote $j = 0$ and the associated utility function as:

$$u_{i0t} = \varepsilon_{i0t}, \tag{2}$$

where x_{jt} is a vector with observed product characteristics including brand intercepts, feature and display variables, p_{jt} is the unit price for brand j at time t, β_i and α_i are individual-level response coefficients, ξ_{jt} is an unobserved demand shock for brand j at time period t, and ε_{ijt} is the unobserved error term that is assumed to be uncorrelated with price. We make assumptions on the error terms and response coefficients as the following:

$$\varepsilon_{ijt} \sim \text{Extreme value } (0, 1), \tag{3}$$
$$\xi_t \sim \text{ MVN } (0, \Sigma_d), \tag{4}$$
$$\theta_i = (\alpha_i, \beta_i')' \sim \text{ MVN } (\overline{\theta}, \Sigma_\theta). \tag{5}$$

The type I extreme value specification of ε_{ijt} leads to a standard logit choice probability for person i choosing brand j at time t,

$$\Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(V_{ijt})}{1 + \sum_k \exp(V_{ikt})}, \tag{6}$$

where

$$V_{ijt} = \beta_i' x_{jt} + \alpha_i p_{jt} + \xi_{jt}. \tag{7}$$

Assuming the sample is representative of the market and households do not make multiple purchases, we obtain market share for brand j at time t as,

$$s_{jt} = \sum_i \frac{s_{ijt}}{I}. \tag{8}$$

If firms use expected demand (s_{jt}) to set price, then price is not exogenously determined. Price and the demand-side error, ξ_t in equation (7), will be correlated because expected demand, used to set prices, is a function of both. That is, it is not possible to write the joint distribution of prices and demand as a conditional demand distribution and a marginal price distribution. Demand and prices are both functions of consumer price sensitivity and other model parameters. Not accounting for the endogenous nature of price will produce biased estimates of model parameters, including household price sensitivity.

2.2. *Supply model—profit maximizing prices*

We assume that each manufacturer maximizes the following objective function

$$\max_{w_i} \pi_i = Ms_i(w_i - c_{mi}) + \phi \sum_{j \neq i} Ms_j(w_j - c_{mj}), \tag{9}$$

where M is the potential market size, w is the wholesale price, c_{mi} is manufacturer i's marginal cost, and $\phi = 1$ or 0. If $\phi = 1$, then manufacturers are involved in tacit collusion. If $\phi = 0$, then they are in Bertrand competition. The first order condition for the manufacturers implies,

$$w - c_m = (H'Q')^{-1}(-s), \tag{10}$$

where $H_{ii} = \partial s_i/\partial p_i, H_{ik} = \phi \partial s_i/\partial p_k$ and $Q_{ik} = \partial p_i/\partial w_k$ $(i = 1, \ldots, J,$ and $k = I, \ldots, J)$.

Next, we turn to the retailer's pricing strategy. For the purpose of illustration, we only model a single retailer's pricing behavior even though competition among multiple retailers is possible. One simple rule the retailer can use is to simply charge a fixed markup over wholesale price for each brand, resulting in the following specification,

$$p_i = w_i + m_i, \tag{11}$$

where m stands for the fixed markup. This pricing strategy infers that $\partial p_i/\partial w_i = 1$ and $\partial p_i/\partial w_j = 0$. Substituting those two conditions into equation (10), we obtain the following pricing equation,

$$p = c_m + m - (H')^{-1} s. \tag{12}$$

The retailer can also use some more sophisticated pricing strategies such as maximizing each brand's profit or jointly maximizing the category profit in a manufacturer Stackelberg game. In such case, the retailer's objective function is given as following,

$$\max_{p_i} \Pi_i = Ms_i(p_i - w_i - c_{ri}) + \theta \sum_{j \neq i} Ms_j(p_j - w_j - c_{rj}), \tag{13}$$

where c_{ri} is the retailers' marginal cost for brand i, and $\theta = 1$ or 0. If $\theta = 0$, then the retailer maximizes profit from each brand separately. Otherwise, the retailer maximizes category profit. The first order condition implies:

$$p - w - c_r = \left(\frac{\partial G'}{\partial p}\right)^{-1} (-s), \tag{14}$$

where $\partial G_i/\partial p_i = \partial s_i/\partial p_i$ and $\partial G_i/\partial p_k = \theta \partial s_i/\partial p_k$ $(i = 1, \ldots, J,$ and $k = 1, \ldots, J)$. Differentiating both sides of equation (14) with respect to w_j, we obtain,

$$\frac{\partial p}{\partial w} = U^{-1} \left(\frac{\partial G}{\partial p}\right), \tag{15}$$

where

$$U_{ik} = \frac{\partial s_i}{\partial p_k} + \frac{\partial G_k}{\partial p_i} + \sum_{l=1}^{J} \frac{\partial^2 G_l}{\partial p_i \partial p_k}(p_l - w_l - c_{rl})$$
$$= \frac{\partial s_i}{\partial p_k} + \frac{\partial G_k}{\partial p_i} + \left(\frac{\partial^2 G}{\partial p_i \partial p_k}\right)\left(\frac{\partial G}{\partial p}\right)^{-1}(-s). \tag{16}$$

The last equality in equation (16) is a result of equation (14). Therefore, combining equations (10), (14) and (15) gives the pricing equations:

$$p = c_m + c_r + \left\{\left[HU^{-1}\left(\frac{\partial G'}{\partial p}\right)\right]\right\}^{-1}(-s) + \left(\frac{\partial G'}{\partial p}\right)^{-1}(-s), \tag{17}$$

where p, c_m, c_r, s and η are $J \times 1$ vectors, and $H, U, \partial G/\partial p$ are $J \times J$ matrix. A detailed expression on the building blocks of the aforementioned matrices is outlined in Appendix A. It is worth noting that the last two terms of equation (17) are profit margins for manufacturers and retailer respectively. Finally, we can specify the manufacturer and retailer cost $c_m + c_r$ as a brand-specific linear function of cost shifters Z, that is

$$c_{mt} + c_{rt} = Z_t'\delta_j + \eta_t, \tag{18}$$

where η is the supply side error which we assume a multivariate normal distribution, that is, $\eta_t \sim MVN(0, \Sigma_s)$, or

$$p_t = Z_t'\delta_j + \left\{\left[HU^{-1}\left(\frac{\partial G'}{\partial p}\right)\right]\right\}^{-1}(-s) + \left(\frac{\partial G'}{\partial p}\right)^{-1}(-s) + \eta_t. \tag{19}$$

The distribution of observed prices is obtained from the distribution of the supply side error by using change-of-variable calculus. This distribution is of non-standard form because equation (19) is implicit in price—i.e., price appears on both the left and right side of the equal sign in the terms H, U, $\partial G/\partial p$ and s. The distribution of observed prices is obtained by defining a new variable $r = p + f(p) - Z\delta$ that is distributed normal with mean 0 and covariance Σ_s. The likelihood for price is obtained in the standard way as the likelihood for r multiplied by the determinant of the Jacobian ($J = \{\partial r_i/\partial p_j\}$). Details of deriving the distribution of prices for parameter estimation are provided below and in Appendix B.

We realize that pricing equations other than what we specified above can be derived as well using alternative competitive structures (Sudhir, 2001b; Villas-Boas, 2002). Our proposed estimation method can be easily extended to those scenarios even though we choose to focus on two cases (Bertrand manufacturers-fixed markup retailer; Bertrand manufacturers-category profit maximizing retailer) for the purpose of method demonstration.

2.3. Supply model—instrumental variable specification

Researchers often prefer a limited information approach because it does not require specification of the supply-side relationship. Instead, it assumes that the observed supply-side variables, such as prices, are stochastic with an error term that is correlated with the demand-side error. The non-zero covariance between the demand and supply errors are the common parameters that create the endogenous relationship. The predicted supply variables are orthogonal to the supply-side error, and, so long as the instruments are also orthogonal, the predicted values can be viewed as exogenous covariates in the demand-side model. While not using the supply equation reduces the information available for estimating the parameters of the demand model, it does not result in inconsistent estimates.

An alternative interpretation of the limited information approach is that it replaces the true supply-side model with a linear model, with instruments as the covariates and error that is correlated with the demand-side error. The common parameters that give rise to simultaneous demand and supply are the covariance parameters among the demand- and supply-side errors. The advantage of this interpretation is that it allows us to compare alternative model specifications within a common model framework. It also emphasizes our view that the analysis of simultaneous demand and supply systems requires a specification of demand and supply relationships. Claims that limited information methods are preferred because they do not require specification of a supply-side model are inconsistent with the goal of analyzing simultaneous demand and supply.

Widely used price instruments are lagged price, lagged share, cost, price in other markets, etc. Since lagged prices are easily available set of instruments for price, we specify the following supply equation:

$$p_{j,t} = \delta_{j0} + \delta_{j1} p_{j,t-1} + \eta_{j,t} \tag{20}$$

and the variance covariance between demand common shock and price equation error as

$$\text{cov}\begin{pmatrix} \xi_{jt} \\ \eta_{jt} \end{pmatrix} = \begin{pmatrix} \Sigma_d & \Sigma'_{ds} \\ \Sigma_{ds} & \Sigma_s \end{pmatrix}, \tag{21}$$

Σ_{ds} will not be equal to zero if price endogeneity is present. We note that lagged price may not be appropriate instruments since it could be correlated with the utility equation error terms due to reasons such as forward buying and stockpiling.

However, Villas-Boas and Winer (1999) show that such correlation under-estimates the effects of endogeneity.

2.4. Bayesian estimation

The technique of data augmentation (Tanner and Wong, 1987) is used to facilitate estimation of the model. Data augmentation involves adding latent variables to the model parameters, and using them as conditioning arguments to simplify estimation. We introduce household specific coefficients $\{\theta_i\}$ and supply shock realizations $\{\xi_t\}$ as augmented, latent variables. Analysis proceeds by iteratively sampling from the full conditional distributions of all model parameters, including the augmented variables. The dependent variables are choice (y_{it}) and prices (p_t), and the model can be written in hierarchical form:

$y_{it}\|p_t, \theta_i, \xi_t, \varepsilon_{it}$	Observed demand,	(22)
$p_t\|\{\theta_i\}, \{\xi_t\}, \delta, \eta_t$	Observed prices,	(23)
$\theta_i\|\overline{\theta}, \Sigma_\theta$	Heterogeneity,	(24)
$\xi_t\|\Sigma_d$	Demand shock,	(25)
$\eta_t\|\Sigma_s$	Supply shock,	(26)
ε_{it}	Extreme value (logit) error,	(27)

where $\theta_i = (\alpha_i', \beta_i')'$. Observed demand for the ith household is dependent on the household's coefficients (θ_i), the demand shock (ξ_t), the unobserved error (ε_{it}) and the explanatory variables, including prices (equation (6)). Observed prices are determined by the set of household coefficients $\{\theta_i\}$, cost shifter coefficients (δ) and the supply shock (η_t), and are set in response to the expected demand across the heterogeneous households (equation (19)). Household coefficients are specified as random-effects, and the demand and supply shocks specified as normally distributed. The demand and supply shocks are assumed to be independent in the full information model, with endogeneity resulting from the common parameters (θ_i). In the limited information model, we must allow for nonzero covariance in the demand and supply shocks and introduce the additional parameter Σ_{ds}.

Given the household coefficients $\{\theta_i\}$ and demand shocks $\{\xi_t\}$, the joint distribution of demand and prices is obtained by multiplying the conditional (on prices) demand density by the marginal price density. The marginal price density, given $\{\theta_i\}$ and $\{\xi_t\}$, is derived from the supply-side error term, η_t. The conditional demand density, given $\{\theta_i\}$, $\{\xi_t\}$ and prices, is multinomial with logit probabilities (see equation (6)). The joint density of all model parameters is then:

$$f(\{\theta_i\},\{\xi_t\},\delta,\overline{\theta},\Sigma_d,\Sigma_s,\Sigma_\theta \mid \{y_{it}\},\{p_t\})$$
$$\propto \prod_{t=1}^{T}\prod_{i=1}^{I}\text{prob}(y_{it}\mid p_t,\theta_i,\xi_t)\pi_1(\xi_t\mid\Sigma_d)\pi_2(p_t\mid\{\theta_i\},\{\xi_t\},\delta,\Sigma_s)\pi_3(\theta_i\mid\overline{\theta},\Sigma_\theta)\pi_4(\delta,\overline{\theta},\Sigma_d,\Sigma_s,\Sigma_\theta), \tag{28}$$

where $\text{prob}(y_{it}|\theta_i,\xi_t)$ is the logit choice probability for household i at time t, π_1 is the density contribution of the demand error ξ_t, π_2 is the density contribution of the observed prices at time t that depend on consumer preferences and price sensitivities $\{\theta_i\}$, demand errors $\{\xi_t\}$, cost variables and coefficients (Z, δ) and the supply-side error (η_t), π_3 is the distribution of heterogeneity and π_4 is the prior distribution on the hyper-parameters.

Estimation is carried out using a Markov chain Monte Carlo procedure that involves generating a sequence of draws from the full conditional distributions of the model (see Gelfand and Smith, 1990; Gelfand et al., 1990). The introduction of the augmented latent variables results in substantial simplification in estimation, including the implementation of change-of-variable calculus needed to derive the likelihood for prices from the supply-side errors. Numerical derivatives are employed to evaluate the Jacobian during estimation. Appendix B provides the conditional distributions for the full information model, and Appendix C provides conditional distributions for the limited information estimation using instrumental variables.

The difference between Bayesian and non-Bayesian analysis of the joint likelihood for choices (y_{it}) and prices (p_t) in equation (28) is two-fold. First, Bayesian analysis does not require the integration of equation (28) with respect to the random-effects (θ_i) and demand shocks (ξ_t) to obtain the marginalized, or unconditional likelihood. Such marginalization is difficult to evaluate because the integral is of high dimension and involves highly non-linear functions of the model parameters, including the Jacobian needed to obtain the distribution of observed prices. An advantage of using a Markov chain Monte Carlo estimator is that the "grid" for conducting numerical integration is determined automatically. Second, a Bayesian analysis requires the presence of a proper prior distribution, $\pi_4(\cdot)$, to ensure the existence of posterior distribution. In our analysis, we specify the prior as proper but relatively diffuse so that it exerts minimal influence on the analysis.

Latitude exists in specifying the nature of the endogenous relationship that reflects the information set used by manufacturers and retailers for setting prices. In our model, the set of household preferences and sensitivity to price, display and feature advertising are assumed to be associated with observed demand and prices. That is, we assume that manufacturers and retailers have access to scanner panel data similar to that used in our analysis, and that these firms use disaggregate estimates of individual demands to guide pricing decisions. The wide-spread use of choice simulators by firms, in which pricing scenarios are explored via spreadsheets, makes this a reasonable assumption. An alternative specification assumes that prices are determined from knowledge of aggregate. This implies that the conditional distribution of prices in equation (23) is a function of the hyper-parameters $\overline{\theta}$ and Σ_θ, instead of $\{\theta_i\}$.

In our analysis we find support for the endogenous specification that assumes manufacturers and retailers set prices based on information about individual demands. The information set used by firms to set prices is an important area for future research, and in our analysis we find that the model based on knowledge of individual demands fits the data better than one that assumes that prices are set based on knowledge of aggregate demand. We note that the posterior distribution of individual demand parameters need not follow a normal distribution because θ_i appears in multiple equations (22)–(24). Hence, assuming that prices are set based on $\{\theta_i\}$ can be thought of as a non-parametric approach to dealing with the distribution of preferences and sensitivities. Operationally, this approach allows us to integrate over the distribution of heterogeneity to obtain choice shares and other elements in the pricing equation (equation (19)) by summing over functions of the augmented latent variables that are available as a by-product of the MCMC estimator.

3. Empirical application

Data were collected by the AC Nielsen Corporation from the Saint Petersburg, Florida, market between January 1997 and September 1998, a total of 92 weeks. Data are from one retail chain, and the product category is domestic light beer, and our analysis focuses on the three major brands: Miller Lite, Bud Light and Coors Light. These brands account for more than 80% share of the total light beer market. We examine demand and prices for the most popular size/form combination—twelve packs of 12 ounce cans. The dataset comprises 185 households who shopped at one store, making at least one purchase during the time period. On average, the households were observed to purchase beer three to four times, for a total of 631 purchase occasions over all households.

Descriptive statistics are reported in Table 1. The data reveal the following characteristics. Coors Light has the highest choice share at 0.41 followed by Bud Light at 0.33 and Miller Lite at 0.26. We also note that Coors Light engages in more feature and display activities than either Miller or Bud, and has the lowest unit price among the three brands.

Table 1. Sample statistics.

Brands	Choice shares	Price ($)	Feature freq.	Display freq.
Miller Lite	0.26	7.46	0.28	0.23
Bud Light	0.33	7.52	0.27	0.24
Coors Light	0.41	7.34	0.36	0.26

Estimation was carried out using the Markov chain Monte Carlo methods outlined in Appendix B and C. Draws from the posterior distributions were used to evaluate means and standard deviations of the parameter estimates. The chain ran for 20,000 iterations and the last 10,000 iterations were used to obtain parameter estimates. Convergence was assessed by starting the chain from multiple points and inspecting time-series plots of model parameters.

Table 2 reports the fit statistics for 10 different models. The model fit is assessed using the marginal density of the data, which is used in Bayesian analyzes to choose among candidate models. The marginal density of the observed choices (y_{it}) and prices (p_t) is computed using the importance sampling method of Newton and Raftery (1994, p. 21) that re-weights the conditional likelihood of the data. Computing the marginal density of the joint data allows us to evaluate the fit of both full and limited-information model specifications.

The first five models ignore heterogeneous consumer responses to brand and merchandizing variables. Model 1 is a standard multinomial logit model where price is treated as exogenous. Model 2 introduces the demand shock, ξ_t, into the model, similar to the analysis of Allenby and Lenk (1994), except the shock is not person-specific. Model 3 assumes Bertrand manufacturers and a fixed markup retailer on the supply side. Model 4 incorporates an alternative competition structure with Bertrand

Table 2. Comparison of model fit.

Model	Demand shock (ξ)	Consumer heterogeneity	Accounting for endogeneity	Retailer pricing strategy	Manufacturer pricing strategy	Log-marginal density (*)
1	No	No	No	N.A.	N.A.	– 3460.91
2	Yes	No	No	N.A.	N.A.	– 3399.96
3	Yes	No	Full information approach	Fixed markup	Maximize own profit	– 3407.79
4	Yes	No	Full information approach	Maximize category profit	Maximize own profit	– 3410.31
5	Yes	No	Limited information approach	$p_{j,t} = \delta_{j0} + \delta_{j1} p_{j,t-1} + \eta_{j,t}$		– 3400.59
6	No	Yes	No	N. A.	N. A.	– 2579.12
7	Yes	Yes	No	N. A.	N. A.	– 2466.98
8	Yes	Yes	Full information approach	Fixed markup	Maximize own profit	– 2389.29
9	Yes	Yes	Full information approach	Maximize category profit	Maximize own profit	– 2377.20
10	Yes	Yes	Limited information approach	$p_{j,t} = \delta_{j0} + \delta_{j1} p_{j,t-1} + \eta_{j,t}$		– 2464.00

(*) Model fit is measured by the log marginal density calculated using the importance sampling method of Newton and Raftery (1994, p. 21).

manufacturers and a category profit-maximizing retailer. Model 5 takes a limited information approach by assuming a pricing equation with last period's price as an instrument. Models 6–10 are the counterparts of Model 1–5 respectively, with household heterogeneity included in the model specification. The fit statistics reflect contribution from observed demand and prices. For the demand-side models (models 1, 2, 6 and 7), prices are assumed to be distributed multivariate normal with independent parameters.

The model fit statistics indicate the following. First, the introduction of the demand shock, ξ_t, leads to improvement in the model at both the aggregate and disaggregate level. The standard logit model (models 1 and 6) has a restrictive IIA (independence of irrelevant alternatives) property that is relaxed when the additional error is added to the model. Second, consumer heterogeneity is an important element of the model specification, associated with a 30% improvement in the log marginal density of models 6–10 relative to models 1–5. Third, the two full information models fit the data better when heterogeneity is present (models 8 and 9 relative to 7 and 10), but fit the data worse when heterogeneity is not present (models 3 and 4 relative to 2 and 5) in the model specification. An attraction of using limited information methods is that the supply side model does not need to be specified. Our results indicate that mis-specification of heterogeneity can mask the importance of the supply side model, as well as bias the inferred nature of competition. In fact, the best fitting aggregate model assumes that prices are exogenous. Finally, the results indicate that model 9 fits the data best—the retailer is more likely to be engaged in maximizing category profits as opposed to fixed markup pricing.

Parameter estimates for the 10 models are presented in Table 3. The posterior mean and standard deviation of the parameters are reported for models 1–5. Estimates of the hyper-parameters of the random-effects distribution are reported for models 6–10. Consistent with previous findings, we find that accounting for price endogeneity using either full information or limited information approach significantly increases the estimated price sensitivity in models without consumer heterogeneity (Villas-Boas and Winer, 1999). Estimates of price sensitivity increase in magnitude from $-$ 0.299 in model 2, to $-$ 0.674 and $-$ 0.838 in models 3 and 4. However, when consumer heterogeneity is included in the model specification in models 6–10, the effect of accounting for price endogeneity has virtually no affect. The estimate of the price sensitivity hyper-parameter decreases slightly from -1.121 in model 6 to -1.076 in model 9, the best fitting model. The results provide evidence that the reported downward bias of price sensitivity does not necessarily exist when consumer heterogeneity is included in the model.

The other coefficients reported in Table 3 display the following pattern. In general, the brand intercept for Coors Light is largest and the brand intercept for Miller Lite is smallest, consistent with the choice shares reported in Table 1. The specific values of the brand intercepts are dependent on the estimated price coefficient (α_p), with more negative values associated with price coefficients closer to zero. The estimates of the feature and display coefficients, and the supply-side coefficients, are roughly equivalent across model specifications.

Table 3. Posterior mean (standard deviation) of model coefficients.

Coefficient	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10
$\overline{\beta}_{\text{Miller}}$	− 3.622 (0.597)	− 2.658 (1.563)	0.223 (1.002)	1.490 (0.827)	− 2.691 (1.741)	0.592 (0.376)	0.516 (0.373)	− 0.607 (0.357)	0.397 (0.488)	0.585 (0.303)
$\overline{\beta}_{\text{Bud}}$	− 3.377 (0.599)	− 2.383 (0.575)	0.480 (0.998)	1.765 (0.823)	− 2.398 (1.753)	1.414 (0.313)	1.779 (0.311)	0.303 (0.392)	1.220 (0.303)	1.440 (0.261)
$\overline{\beta}_{\text{Coors}}$	− 3.236 (0.589)	− 2.416 (1.533)	0.458 (0.939)	1.738 (0.759)	− 2.421 (1.725)	1.608 (0.288)	1.801 (0.234)	0.252 (0.350)	1.530 (0.302)	1.464 (0.302)
$\overline{\beta}_{f}$	0.303 (0.055)	0.277 (0.167)	0.025 (0.124)	− 0.067 (0.140)	0.281 (0.152)	− 0.128 (0.132)	− 0.003 (0.210)	0.170 (0.146)	− 0.058 (0.162)	0.145 (0.188)
$\overline{\beta}_{d}$	0.307 (0.049)	0.194 (0.130)	0.173 (0.175)	0.114 (0.161)	0.195 (0.145)	0.199 (0.118)	0.060 (0.212)	0.074 (0.174)	0.155 (0.152)	− 0.070 (0.172)
$\overline{\alpha}_{p}$	− 0.156 (0.078)	− 0.299 (0.205)	− 0.674 (0.129)	− 0.838 (0.108)	− 0.301 (0.229)	− 1.050 (0.045)	− 1.121 (0.049)	− 0.944 (0.050)	− 1.076 (0.046)	− 1.095 (0.036)
$\delta_{\text{Miller},0}$	N.A.	N.A.	5.914 (0.298)	4.935 (0.324)	4.476 (1.013)	N.A.	N.A.	6.070 (0.107)	4.765 (0.158)	4.496 (1.061)
$\delta_{\text{Bud},0}$	N.A.	N.A.	5.961 (0.302)	4.982 (0.326)	4.421 (0.988)	N.A.	N.A.	5.713 (0.135)	4.892 (0.138)	4.481 (0.989)
$\delta_{\text{Coors},0}$	N.A.	N.A.	5.783 (0.300)	4.805 (0.325)	4.517 (0.988)	N.A.	N.A.	5.757 (0.121)	4.705 (0.153)	4.564 (0.985)
$\delta_{\text{Miller},1}$	N.A.	N.A.	N.A.	N.A.	0.400 (0.135)	N.A.	N.A.	N.A.	N.A.	0.398 (0.142)
$\delta_{\text{Bud},1}$	N.A.	N.A.	N.A.	N.A.	0.412 (0.131)	N.A.	N.A.	N.A.	N.A.	0.405 (0.132)
$\delta_{\text{Coors},1}$	N.A.	N.A.	N.A.	N.A.	0.386 (0.134)	N.A.	N.A.	N.A.	N.A.	0.379 (0.134)

Table 4. Posterior mean (standard deviation) of heterogeneity covariance matrix $\sum_{\theta}$.

Model 9	Miller	Bud	Coors	Feature	Display	Price
Miller	8.114 (1.790)					
Bud	0.879 (0.950)	6.031 (1.720)				
Coors	– 0.680 (0.830)	– 1.159 (0.540)	3.016 (0.881)			
Feature	0.295 (0.463)	– 0.178 (0.389)	1.096 (0.388)	1.350 (0.337)		
Display	0.106 (0.442)	0.329 (0.402)	0.368 (0.328)	0.110 (0.229)	0.952 (0.287)	
Price	– 0.315 (0.113)	– 0.240 (0.128)	– 0.168 (0.094)	– 0.147 (0.055)	– 0.097 (0.054)	0.135 (0.018)

The covariance matrix of the distribution of heterogeneity for Model 9, the best fitting model, is reported in Table 4. The pattern of the heterogeneity distribution is similar across Models 6–10, and therefore we only choose to present the results from Model 9. The estimates indicate that there exists substantial heterogeneity in brand preferences. Heterogeneity is also present in the slope coefficients. The posterior mass of most covariance elements are close to zero.

Tables 5 and 6 report the covariance matrices of the demand and supply shocks for the best fitting model (model 9), and the model that employs a limited-information like specification for prices (model 10). As expected, we find that the shocks are larger in magnitude for the limited information specification (Table 6) than for the full information specification (Table 5). Moreover, we find that the estimated covariance between the demand and supply shocks in model 10 all have mass close to zero. The absence of estimates with posterior mass away from zero implies that prices are exogenous. The limited information approach therefore leads

Table 5. Covariance matrix of demand and supply shocks (model 9).

	Demand shocks (ξ_t)		
	Miller	Bud	Coors
Miller	0.361 (0.097)		
Bud	– 0.012 (0.056)	0.275 (0.069)	
Coors	– 0.040 (0.095)	– 0.008 (0.076)	0.692 (0.181)
	Supply shocks (η_t)		
	Miller	Bud	Coors
Miller	0.157 (0.035)		
Bud	0.029 (0.020)	0.151 (0.030)	
Coors	0.003 (0.014)	0.008 (0.014)	0.107 (0.019)

Table 6. Covariance matrix for demand and supply shocks (Model 10).

	ξ_{Miller}	ξ_{Bud}	ξ_{Coors}	η_{Miller}	η_{Bud}	η_{Coors}
ξ_{Miller}	0.635 (0.153)					
ξ_{Bud}	−0.024 (0.100)	0.521 (0.124)				
ξ_{Coors}	0.008 (0.129)	−0.018 (0.113)	0.965 (0.236)			
η_{Miller}	−0.004 (0.046)	−0.011 (0.038)	−0.027 (0.051)	0.200 (0.029)		
η_{Bud}	0.005 (0.048)	0.036 (0.044)	−0.040 (0.056)	0.008 (0.022)	0.224 (0.033)	
η_{Coors}	−0.028 (0.044)	0.016 (0.041)	0.032 (0.057)	−0.011 (0.022)	0.004 (0.024)	0.219 (0.034)

to incorrect inferences about prices since model 9, the best fitting model, supports the assumption that prices are endogenously determined.

Finally, the use of MCMC estimation with data augmentation facilitates inference about all model parameters, including the demand shocks, ξ_t. Figure 1 displays plots of the posterior mean of ξ_t against prices for each brand in the analysis. Each point in the plots corresponds to the weekly price of the offering and the associated expected value of the realized demand shock, given the data. The expectation is trivial to compute because realizations of ξ_t are generated at each iteration of the Markov chain, and the mean of the posterior distribution of ξ_t is obtained simply by averaging over the values of these realizations. The plots indicate that the demand shocks are positively correlated with prices, leading to the improved fit statistics for model 9 where prices are assumed endogenous, relative to model 7, the corresponding exogenous specification. Demand shocks are largest when prices are high, suggesting that there exist omitted variables in the analysis, such as media advertising, that are negatively correlated with the timing of price promotions. However, despite the presence of price endogeneity, we find that there exists little difference in the estimated consumer price sensitivity.

4. Summary and concluding remarks

Consumer choice modeling has been an important research area in marketing over the past 20 years. The availability of household panel data enables marketing academicians and practitioners to gain a much richer and accurate understanding on the origins of consumer preferences, and consumer sensitivity to marketing variables such as promotion and price. A stream of research within the choice modeling literature has identified the importance of respondent heterogeneity (Allenby and Rossi, 1999). Another has identified the importance of accounting for possible endogeneity of variables, such as price, that are set by sellers in anticipation of marketplace response (Berry, 1994; Villas-Boas and Winer, 1999). In this paper, we propose a method of incorporating heterogeneity and supply-side effects in a model of consumer demand.

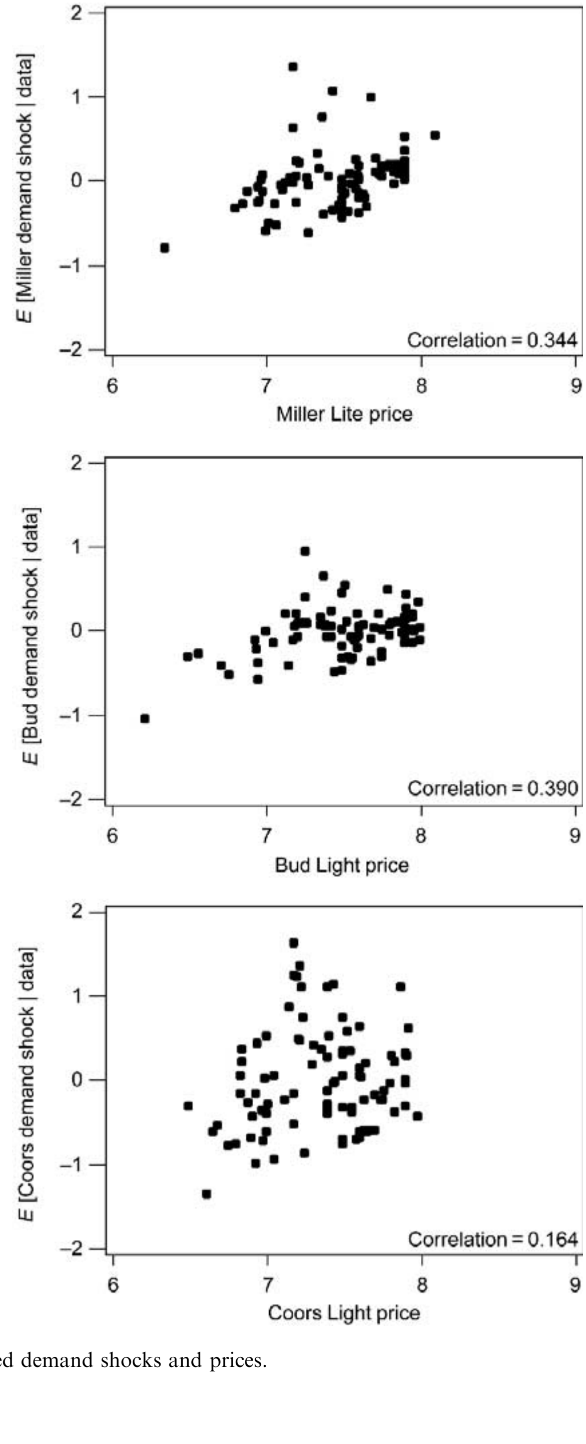

Figure 1. Correlated demand shocks and prices.

Computational challenges arise in the estimation of simultaneous demand and supply models when heterogeneity is incorporated in a random-effects specification. In a classic (frequentist) statistical analysis, the random-effects and demand shocks are integrated from the likelihood, and this task can be burdensome in non-linear models of consumer choice. Methods have been proposed for dealing with this challenge using limited information analysis with instrumental variables. However, these procedures can be inaccurate because of the reduced efficiency associated with limited information analysis, and the need to identify suitable instruments used in the analysis.

The proposed Bayesian approach has known finite-sample properties and deals directly with the complicated likelihood function in simultaneous demand and supply models by employing the technique of data augmentation. Data augmentation leads to less computational burden because it introduces latent variables that are used as conditioning arguments in the Markov chain Monte Carlo analysis. The approach is applied to a data set of light beer purchases, using two alternatives to account for possible price endogeneity—full information and limited information specifications. The analysis demonstrates the importance of incorporating heterogeneity into the model when assessing the presence of endogenous effects, and the difference between full information and limited information analysis. Consistent with previous research, we find that the price coefficient is more negative when endogeneity is incorporated in the no-heterogeneity case. However, when heterogeneity is present, the effect of endogeneity is much smaller. This result is persistent across the alternative structures of competition.

We also find that the consumer heterogeneity specification is important in determining the appropriateness of a full information approach over a limited information approach. A criticism of using a full information approach is that it relies on specific assumptions about the supply-side structure—the game governing how firms compete with each other. The use of instruments to account for price endogeneity is often seen as less risky because it requires fewer assumptions. Our results show that the instrumental variable approach performs worse than the full information approach under the two competition structures when consumer heterogeneity is present in the model. In contrast, when consumer heterogeneity is not captured in the demand side specification, the instrumental variable-type specification yields a model fit similar to the full information specification.

There are numerous applications of our method in studying simultaneous systems of demand and supply in marketing. Management rarely conducts experiments in which resources are allocated randomly. Instead, resources are allocated to obtain the greatest return. Prices are set, products are offered, communication strategies designed, and distribution channels selected to maximize the likelihood of success, with management anticipating the reactions of consumers and competitors. Our approach can handle non-linear model structures and yields exact finite sample inference. We therefore anticipate application of the method to the evaluation of resource allocation by firms, and the study of systems in general.

Appendix A: Calculation of the building blocks for matrices used in supply side specification

$$\frac{\partial s_j}{\partial p_j} = \frac{1}{I}\sum_{i=1}^{I} \alpha_i s_{ij}(1 - s_{ij}), \qquad \frac{\partial s_j}{\partial p_k} = -\frac{1}{I}\sum_{i=1}^{I} \alpha_i s_{ij} s_{ik}$$

$$H_{jj} = \frac{1}{I}\sum_{i=1}^{I} \alpha_i s_{ij}(1 - s_{ij}), \qquad H_{jk} = -\frac{\phi}{I}\sum_{i=1}^{I} \alpha_i s_{ij} s_{ik}$$

$$\frac{\partial G_j}{\partial p_j} = \frac{1}{I}\sum_{i=1}^{I} \alpha_i s_{ij}(1 - s_{ij}), \qquad \frac{\partial G_j}{\partial p_k} = -\frac{\theta}{I}\sum_{i=1}^{I} \alpha_i s_{ij} s_{ik}$$

$$\frac{\partial^2 G_h}{\partial p_k \partial p_j} = \frac{-\theta/I\partial\left(\sum_{i=1}^{I} \alpha_i s_{ih} s_{ik}\right)}{\partial p_j} = \frac{2\theta}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ih} s_{ij} s_{ik}$$

$$\frac{\partial^2 G_h}{\partial p_k \partial p_k} = \frac{-\theta/I\partial\left(\sum_{i=1}^{I} \alpha_i s_{ih} s_{ik}\right)}{\partial p_k} = \frac{2\theta}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ih} s_{ik}^2 - \frac{\theta}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ih} s_{ik}$$

$$\frac{\partial^2 G_k}{\partial p_k \partial p_j} = \frac{1/I\partial\left(\sum_{i=1}^{I} \alpha_i s_{ik}(1 - s_{ik})\right)}{\partial p_j} = \frac{2}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ij} s_{ik}^2 - \frac{1}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ij} s_{ik}$$

$$\frac{\partial^2 G_k}{\partial p_j \partial p_k} = \frac{-\theta/I\partial\left(\sum_{i=1}^{I} \alpha_i s_{ik} s_{ij}\right)}{\partial p_k} = \frac{2\theta}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ij} s_{ik}^2 - \frac{\theta}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ij} s_{ik}$$

$$\frac{\partial^2 G_k}{\partial p_k \partial p_k} = \frac{1/I\,\partial\left(\sum_{i=1}^{I} \alpha_i s_{ik}(1 - s_{ik})\right)}{\partial p_k} = \frac{1}{I}\sum_{i=1}^{I} \alpha_i^2 s_{ik}(1 - s_{ik})(1 - 2s_{ik}).$$

.

Appendix B: Markov Chain Monte Carlo estimation (full information approach)

The full information approach assumes a certain competition structure governs the pricing behavior of manufacturers and retailer. Estimation is carried out by sequentially generating draws from the following distributions:

1. Generate ξ_t

$$[\xi_t|^*] \propto \prod_i \Pr(\text{choice}_{it}) \cdot |J_t| \cdot (p_t + f_t - Z_t\delta \sim MVN(0, \Sigma_s)) \cdot (\xi_t \sim MVN(0, \Sigma_d)),$$

where

$$\Pr(\text{choice}_{it} \neq 0) = \prod_{j=1}^{J} \left(\frac{\exp(V_{ijt})}{1 + \sum_{k=1}^{J} \exp(V_{ikt})} \right)^{I(\text{choice}_{ijt}=1)}$$

(choose one of the brands)

$$\Pr(\text{choice}_{it} = 0) = \frac{1}{1 + \exp(V_{it})} \qquad \text{(choose outside good)}$$

$$V_{ijt} = \beta_i' x_{jt} + \alpha_i p_{jt} + \xi_{jt}$$

$$|J_t| \text{is the Jacobian} = \begin{vmatrix} \frac{\partial r_1}{\partial p_1} & \frac{\partial r_1}{\partial p_2} & \frac{\partial r_1}{\partial p_3} \\ \frac{\partial r_2}{\partial p_1} & \frac{\partial r_2}{\partial p_2} & \frac{\partial r_2}{\partial p_3} \\ \frac{\partial r_3}{\partial p_1} & \frac{\partial r_3}{\partial p_2} & \frac{\partial r_3}{\partial p_3} \end{vmatrix}_t \qquad \text{where } r_t = p_t + f(p_t) - Z_t\delta,$$

$-f_{jt}$ is margin for manufacturer j and the retailer at time t. For example, in the case where manufacturers are fully competing and the retailer uses a simple fixed markup strategy, then

$$f_{jt} = \frac{\sum_1^I s_{ijt}}{\sum_{i=1}^{I} \alpha_i s_{ijt}(1 - s_{ijt})},$$

where

$$s_{ijt} = \frac{\exp(V_{ijt})}{1 + \sum_{k=1}^{J} \exp(V_{ikt})}.$$

In the case where manufacturers are fully competing and the retailer maximizes the category product, then

$$f_t = \left\{ \left[H_t U_t^{-1} \left(\frac{\partial G_t'}{\partial p} \right) \right] \right\}^{-1} s_t + \left(\frac{\partial G_t'}{\partial p} \right)^{-1} s_t$$

and $s_t, H_t, U_t, \partial G_t / \partial p$ are defined as before.

A Metropolis-Hastings algorithm with a random walk chain is used to generate draws of ξ_t (see Chib and Greenberg, 1995, p. 330, method 1).

2. Generate $\theta_i = [\beta_i, \alpha_i]$ (for $i = 1, \ldots, I$)

$$[\theta_i|^*] \propto \prod_{t=1}^{T} \Pr(\text{choice}_{it}) \cdot \prod_{t=1}^{T} |J_t| \cdot \prod_{t=1}^{T} (p_t + f_t - Z_t\delta \sim N(0, \Sigma_s)) \cdot \left(\theta_i \sim N(\overline{\theta}, \Sigma_\theta)\right).$$

A Metropolis-Hastings algorithm with a random walk chain is used to generate draws of ξ_t

3. Generate $\overline{\theta}$

$$[\overline{\theta}|^*] = \text{Normal}(A, B),$$

where

$$A = B\left(\frac{\Sigma_\beta}{I}\right)^{-1}\left(\frac{\sum_i \theta_i}{I}\right)$$
$$B = \left(\left(\frac{\Sigma_\theta}{I}\right)^{-1} + V_0^{-1}\right)^{-1}$$
$$V_0 = 100I.$$

4. Generate Σ_θ

$$[\Sigma_\theta|^*] \propto \text{Inverted Wishart}\left(\sum_{t=1}^{I}(\theta_i - \overline{\theta})(\theta_i - \overline{\theta})' + G, I + g\right)$$

where $G = 10I$, and $g = 10$, and I indexes identity matrix with dimension equal to the number of parameters in the mean utility.

5. Generate Σ_d

$$[\Sigma_d|^*] \propto \text{Inverted Wishart}\left(\sum_{t=1}^{T}\xi_t'\xi_t + G, T + g\right),$$

where $G = 5I$, and $g = 5$, and I indexes identity matrix with dimension equal to J.

6. Generate $\delta = (\delta_{10}, \delta_{20}, \delta_{30})$

$$[\delta|^*] = MVN(v, \Omega),$$

where

$$v = \Omega(Z'\Sigma_d^{-1}(p + f) + D^{-1}r_0)$$
$$\Omega = (D^{-1} + Z'\Sigma_d^{-1}Z)^{-1}$$
$$r_0 = (0, 0)'$$
$$D = 100I.$$

7. Generate Σ_s

$$[\Sigma_s|^*] \propto \text{Inverted Wishart}\left(\sum_t (P_t + f_t - Z_t'\delta)'(P_t + f_t - Z_t'\delta) + G, T + g\right),$$

where $G = 5I$, and $g = 5$, and I indexes identity matrix with dimension equal to J.

Appendix C: Markov Chain Monte Carlo estimation (limited information approach)

The limited information approach specifies a reduced form of pricing equation. The example we illustrate here is one within which price is related to last period price in a linear fashion. In specific,

$$p_{j,t} = \delta_{j0} + \delta_{j1} p_{j,t-1} + \eta_{j,t},$$

where

$$\text{cov}\left(\xi_{jt}', \eta_{jt}'\right) = \begin{pmatrix} \Sigma_d & \Sigma_{ds} \\ \Sigma_{ds} & \Sigma_s \end{pmatrix}.$$

The pricing equation can be also written as following, assuming a case of three brands:

$$p_{jt} = W_t\delta + \eta_{jt},$$

where

$$W_t = \begin{pmatrix} 1 & 0 & 0 & p_{t-1,1} & 0 & 0 \\ 0 & 1 & 0 & 0 & p_{t-1,2} & 0 \\ 0 & 0 & 1 & 0 & 0 & p_{t-1,3} \end{pmatrix}$$

$$\delta = (\delta_{10}, \delta_{20}, \delta_{30}, \delta_{11}, \delta_{21}, \delta_{31}).$$

Estimation is carried out by sequentially generating draws from the following distributions:

1. Generate ξ_t

$$[\xi_t \mid {}^*] \propto \prod_i \Pr(\text{choice}_{it}) \cdot \left(\begin{pmatrix} \xi_t \\ p_t - Z_t \delta \end{pmatrix} \sim MVN(0, \Sigma) \right).$$

The choice probability is specified as before.
We use Metropolis-Hastings algorithm with a random walk chain to generate draws of ξ_t.

2. Generate $\theta_i = [\beta_i, \alpha_i]$ (for $i = 1, \ldots, I$)

$$[\theta_i \mid {}^*] \propto \prod_{t=1}^{T} \Pr(\text{choice}_{it}) \cdot \left(\theta_i \sim N(\overline{\theta}, \Sigma_\theta)\right).$$

We use Metropolis-Hastings algorithm with a random walk chain to generate draws of ξ_t.

3. Generate $\overline{\theta}$

$$[\overline{\theta} \mid {}^*] = \text{Normal } (A, B),$$

where

$$A = B\left(\frac{\Sigma_\beta}{I}\right)^{-1} \left(\sum_i \frac{\theta_i}{I}\right)$$
$$B = \left(\left(\frac{\Sigma_\theta}{I}\right)^{-1} + V_0^{-1}\right)^{-1}$$
$$V_0 = 100I.$$

4. Generate Σ_θ

$$[\Sigma_\theta \mid {}^*] \propto \text{Inverted Wishart}\left(\sum_{i=1}^{I} (\theta_i - \overline{\theta})(\theta_i - \overline{\theta})' + G, I + g\right),$$

where $G = 10I$, and $g = 10$, and I indexes identity matrix with dimension equal to the number of parameters in the mean utility.

5. Generate δ

$$[\delta \mid {}^*] = MVN(v, \Omega),$$

where

$$\begin{aligned}
v &= \Omega(Z'\Delta^{-1}(p-f) + D^{-1}r_0)\\
\Omega &= (D^{-1} + Z'\Delta^{-1}Z)^{-1}\\
f &= \Sigma_{ds}\Sigma_d^{-1}\xi\\
\Delta &= \Sigma_s - \Sigma_{ds}\Sigma_d^{-1}\Sigma_{ds}'\\
r_0 &= (0,0)'\\
D &= 100I
\end{aligned}$$

6. Generate Σ

$$[\Sigma \mid {}^*] \propto \text{ Inverted Wishart}\left(\sum_{t=1}^{T}\begin{pmatrix}\xi_t\\ P_t - Z_t\delta\end{pmatrix}'\begin{pmatrix}\xi_t\\ P_t - Z_t\delta\end{pmatrix} + G, T+g\right),$$

where $G = 10I$, and $g = 10$, and I indexes identity matrix with dimension equal to $2J$.

Acknowledgments

The authors thank A.C. Nielsen for providing the data and Mark Garratt from the Miller Brewing Company for providing insightful discussions. The authors thank Peter Rossi, Pradeep Chintagunta and Jean-Pierre Dube for helpful comments.

References

Allenby, G.M. and P.J. Lenk. (1994). "Modeling Household Purchase Behavior with Logistic Normal Regression", Journal of the American Statistical Association 89, 1218–1231.

Allenby, G.M. and P. Rossi. (1999). "Marketing Models of Consumer Heterogeneity", Journal of Econometrics 89, 57–78.

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.

Chib, S. and E. Greenberg. (1995). "Understanding the Metropolis-Hastings Algorithm", The American Statistician 49, 327–335.

Chintagunta, P., J.P. Dube, and K.Y. Goh. (2002). "Targeted Pricing and the Estimation of Consumer Choice Models in the Presence of Unmeasured Product Characteristics", Working paper, University of Chicago.

Draganska, M. and D. Jain. (2002). "Structural Models of Competitive Market Behavior: An Estimation Approach Using Disaggregate Data", Working paper, Stanford University.

Gelfand, A.E. and A.F.M. Smith. (1990). "Sampling-Based Approaches to Calculating Marginal Densities", Journal of the American Statistical Association 85, 398–409.

Gelfand, A.E., S. Hills, A. Racine-Poon and A.F.M. Smith. (1990). "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling", Journal of the American Statistical Association 85, 972–985.

Goolsbee, A. and A. Petrin. (2001). "Consumer Gains from Direct Broadcast satellites and the Competition with Cable TV", Working Paper, The University of Chicago GSB.

Kuksov, D. and M. Villas-Boas. (2001). "Endogeneity and Individual Consumer Choice", Working Paper, UC-Berkeley.

Manchanda, P., P. Chintagunta, and P.E. Rossi. (2003). "Response Modeling with Non-Random Marketing Mix Variables", Working paper, Graduate School of Business, University of Chicago.

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

Newton, M.A. and A.E. Raftery. (1994). "Approximating Bayesian Inference with the Weighted Likelihood Bootstrap", Journal of the Royal Statistical Society (B) 56, 3–48.

Petrin, A. and K. Train. (2002). "Omitted Product Attributes in Discrete Choice Models", Working paper, UC-Berkeley.

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

Sudhir, K. (2001b). "Structural Analysis of Competitive Pricing in the Presence of a Strategic Retailer", Marketing Science 20(3), 244–264.

Tanner, M.A. and W.H. Wong. (1987). "The Calculation of Posterior Distributions by Data Augmentation", Journal of the American Statistical Association 82, 528–550.

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

Villas-Boas, M. and Y. Zhao. (2003). "The Ketchup Marketplace: Retailer, Manufacturers, and Individual Consumers", Working paper, UC-Berkley.

Villas-Boas, S.B. (2002). "Vertical Contracts Between Manufacturers and Retailers: An Empirical Analysis", mimeo, University of California at Berkeley.

Quantitative Marketing and Economics, 1, 277–283, 2003.

Comment

PATRICK BAJARI
Department of Economics, Duke University and NBER

1. Introduction

Yang, Chen and Allenby develop a novel Bayesian approach for estimating a structural model of supply and demand. The framework they develop includes, as a special case, a model similar to Berry, Levinsohn and Pakes (1995, henceforth BLP). For many researchers in Industrial Organization, and to a lesser extent in Marketing, BLP is a benchmark empirical model of a differentiated products market. This framework has been applied, and modified, by numerous researchers including Nevo (2000), Petrin (2002), Davis (2003), Leslie (2003), and Villas-Boas (2003).

Bayesian methods are mysterious to many empirical researchers since they are not commonly used in the literature and most leading PhD programs do not cover Markov Chain Monte Carlo (MCMC) in first year or even advanced econometrics courses. Many researchers will therefore be inclined to ask "Why use a Bayesian approach? Doesn't that involve specifying a prior distribution? Are my results meaningful if they depend on the specification of a subjective distribution of prior beliefs?"

In this discussion, I have two objectives. The first is to provide a bit of background about some possible merits of a Bayesian approach from the perspective of an applied economist. Standard textbooks compare Bayesian and Classical approaches from a theoretical point of view. My comments, however, are from the perspective of an applied researcher who has used Bayes in a few projects. My comments will apply not only to Yang, Chen and Allenby, but also to the use of Bayesian approaches more generally.

The second objective is to discuss some issues that have not been resolved by this paper. The limitations I shall discuss are not unique to this paper and are known to seasoned researchers in Bayesian Econometrics, Industrial Organization and Marketing. However, for those not familiar with these methods, I hope that my comments will put the paper into perspective and suggest some directions for future research.

2. Merits of a Bayesian approach

In my experience, a Bayesian approach has several potential benefits. The first is numerical. In certain problems, Bayesian approaches are easier to program and convergence is more robust than in Classical estimators which depend on minimizing a nonlinear, simulated objective function. Also, a Bayesian approach permits exact finite sample analysis and does not rely on common asymptotic approximations. The second is that Bayesian analysis facilitates rational decision making. In some industry and policy problems, specifying the utility function and accounting for uncertainty about key parameters will assists the researcher in making a sound decision. The third is that the prior distribution can be a useful tool. In industry or policy environments where a decision *must* be made, the available data is typically much less rich than we desire. However, the researcher may have access to industry experts or other sources that can be used to specify a prior distribution. If the data is very limited, the analyst should not ignore useful a priori information since estimates that are agnostic about the prior may not be very informative.

2.1. Benefit #1: Numerical

Over the last decade, Bayesian methods have increasingly been applied in Statistics and Econometrics. Much of the recent interest in Bayes is due to the application of MCMC methods, such as Gibbs sampling, to Bayesian Statistics. (See Chib and Greenberg (1995), Gelman, Rubin and Stern (2003) and Geweke (1997) for excellent surveys of these methods.) In MCMC, given a prior and a likelihood function, the econometrician simulates a Markov chain with an invariant distribution equal to the posterior distribution.

Discrete choice is a particularly attractive area for application of MCMC. The likelihood function generated by a standard discrete choice model is a fairly complicated high dimensional integral, with upper and lower bounds of integration that depend on the latent utilities in an intricate manner. In applied work, researchers have typically used models such as the logit or nested logit that simplify the computation of these integrals. Bayesian methods have increased the set of models that are computationally tractable. See Albert and Chib (1993), McCulloch and Rossi (1994), Geweke, Keane and Runkle (1996), and Geweke, Gowrisankaran and Town (2003) for examples. Similarly, the models considered by Yang, Chen and Allenby would be difficult, perhaps not even possible, to estimate using maximum likelihood.

I have personally found that MCMC has three potentially numerical advantages. The first is that MCMC is often easy to program. The simulations in MCMC require the analysts to draw random numbers from standard distributions and evaluate standard parametric density functions. This is often an easy program to write and the researcher is less likely to make coding mistakes than in estimators that require numerical optimization. Thus, less time is spent on software development and waiting for the estimator to converge.

The second advantage is that in many problems convergence of MCMC is surprisingly robust. My experience with the multinomial probit and auction models (see Bajari and Hortacsu (2003) and Bajari and Ye (2003)) is that even in fairly complicated models, after an initial "burn in" the simulations appear to quickly converge to the posterior. From talking with other applied researchers, my belief is that methods based on numerical optimization frequently involve considerably more trail and error. The optimization routines frequently reach pathological regions of the parameter space where the objective function is ill behaved and many starting points need to be explored. While an experienced practitioner will understand the major pitfalls, less experienced researchers are likely to make mistakes.

A final advantage of MCMC, as noted by the authors, is that it facilitates exact, finite sample analysis. Standard asymptotic approximations are not required to explore the properties of the estimator. While Monte Carlo evidence may be suggestive, it is difficult to know in practice if we have a sufficiently large number of observations for asymptotic approximations to be valid.

2.2. *Benefit #2: Decision theory*

Outside of academia, most end users of statistics want to use their parameter estimates to assist them with decision making. They do not typically estimate the elasticity of product level demand curves out of idle curiosity. Instead, in industry, these elasticities might be used to assist the firm in making a pricing decision. In a policy application, the elasticity estimates might have a bearing on whether an antitrust official decides that a proposed merger will lead to excessive market power.

In our academic lives, if our estimates do not turn out to be very appealing, we can simply choose not to submit our paper to a journal. In industry and policy settings, actors do not have this luxury. Even if there is considerable uncertainty about the parameter values, typically *a decision still must be made*.

In economic theory, a rational agent who is uncertain about the parameters that enter into her utility is typically assumed to act as a Bayesian. MCMC output can be used to quickly evaluate the expected utility from alternative actions. Since it facilitates decision making, the approach of Yang, Chen and Allenby has potential benefits compared to Classical alternatives.

In industry, firms could use this type of analysis to compute the posterior expected profits from alternative pricing, promotions and other marketing variables. A nice example of the use of Bayes this type of decision is Rossi, McCulloch and Allenby (1996). In policy applications, such as antitrust, we could evaluate the expected costs and benefits of a proposed merger. Clearly, our tools are not yet perfected for these types of applications. However, they are promising and may eventually offer a more rigorous alternative to current practice.

2.3. *Benefit #3: Priors can be useful*

Many econometricians are well versed in how to flexibly estimate models using, for instance, non-parametric approaches. Unfortunately, in many applications, the available data is often quite limited and therefore very flexible estimators may not be very useful. This is particularly important in differentiated products markets where the number of own and cross price elasticities is equal to the number of products squared. Even in very large scanner data sets, there will not be enough data to estimate all of these elasticities in a completely flexible manner.

What should we do if there is not enough data to flexibly estimate our desired model? One alternative is to simply throw up our hands and ignore the data because it is not sufficiently informative without making ad hoc assumptions. However, one tool to increase the informativeness of the available data is to carefully specify a prior distribution.

The models considered by Yang, Chen and Allenby are almost always applied in a single industry. In the course of conducting such research, we have the chance to meet industry participants whose livelihoods depend on understanding these markets. In antitrust problems, there are opportunities for the antitrust official to speak to industry officials and review industry documents at great length. If the researcher is creative, he can elicit information from "industry experts" to form a prior. The data then can be used to update the beliefs of a well informed industry expert, instead of ignoring the available information contained in the data. If we do not use the data, the beliefs of some industry expert, without updating, may likely determine the decision.

In Bajari and Ye (2003), we elicited the beliefs of two industry experts about the construction industry. These industry experts had 50 years of combined experience. Not surprisingly, they had a much better understanding of how firms compete in this industry than we did. Also, these experts, since they operated large construction companies, spent considerable effort to learn about their own and their competitors' costs.

We elicited a distribution of markups from these industry experts and used it to induce a prior distribution over structural costs parameters. While these experts did not know the markups of their competitors exactly, they felt confident in supplying fairly tight bounds. This informed our posterior distribution of costs parameters and helped us to test between alternative models of industry equilibrium (e.g. competition versus collusion).

In differentiated products markets, industry experts might have useful information about which characteristics are most valued by consumers, the markups associated with various products and which products compete most intensely with each other. When the data is very limited, this information might be useful in specifying a prior distribution over structural demand and cost parameters.

In some cases, the views of industry experts may be quite inaccurate. However, as applied researchers, when we reach conclusions that are wildly inconsistent with industry experts, it is often due to our own ignorance about the industry. Eliciting priors is therefore a useful tool for us in assessing the plausibility of our results. At a minimum, we will be aware if our results are within the support of the priors of industry experts.

3. Unresolved problems

The Bayesian framework in Yang, Chen and Allenby has a number of attractive features compared to Classical alternatives. However, there are still many unresolved issues in estimating supply and demand in differentiated products markets. In this section, I will briefly discuss three limitations of this framework. Many of the limitations that I will discuss are shared by all papers in the literature. However, I hope that these comments might be useful to other researchers.

3.1 Problem #1: Numerical issues

While MCMC has computational advantages, there are still some potential pitfalls. The first is assessing convergence of the posterior can be difficult. The invariant distribution of the Gibbs sampler is equal to the posterior. However, it is not always clear in practice to determine how many draws are required to converge to the posterior.

The second is that the simulations may depend heavily on the starting point used in the Gibbs sampler. If the likelihood function is a multimodal disaster, the Gibbs sampler could remain trapped in one region of the parameter space. Gibbs samplers are subject to many of the criticisms that are made of non-linear optimization. That is, we may not know whether we are learning about the posterior globally or merely a potentially pathological local region.

3.2. Problem #2: Do games generate a well defined likelihood?

Few games generate a unique equilibrium. While some games used in empirical analysis, such as auctions have a unique equilibrium, many others do not. To the best of my knowledge, there are no primitive conditions that guarantee that the pricing games studied in this paper have a unique equilibrium.

Given a fixed value for all of the other parameters and a fixed set of covariates, what is the likelihood function if there are two or more equilibrium? How would we assign probabilities to the endogenous variables? What would we do if there were no equilibrium? Recently, researchers have become increasingly aware of the importance of this problem. (For instance, see Tamer (2003) for an approach to estimating entry games with multiple equilibrium.)

If the researcher does not specify a model of how alternative equilibrium are selected, it is not clear how to write the likelihood function for oligopoly games. Therefore, the posterior is not well defined. I do not believe that Yang, Chen and Allenby addressed this problem.

I conjecture that the authors implicitly solve this problem by conditioning on the equilibrium that they observe. That is, the posterior is conditional on the values of the endogenous prices and quantities seen in the data. This is what occurs in many other estimators of games, such as BLP (1995), Aguirrebagiria and Mira (2003), Pesendorfer and Schmidt-Dengler (2003) and Pakes, Ostrovsky and Berry (2003). However, the formal conditions under which this conditioning is valid are not stated in Yang, Chen and Allenby. Imagine, for instance, that there are two equilibria, A and B. Suppose on even numbered months, A is played and on odd numbered months B is played. Alternatively, A could always be played or B could always be played. Would our parameter estimates be invariant to these alternative assumptions? I conjecture that the answer is no.

3.3. Problem #3: Other endogeneity problems

In this paper, endogenous prices and quantities are the main concern. However, there are other potential endogeneity problems. For instance, the "demand shocks" considered in this paper can be interpreted as product characteristics that are observed to the consumer, but not by the economist. In many industries, this may be the most reasonable interpretation of the error term.

In this case, the identifying assumptions could be interpreted as the unobserved product characteristics are independent of observed product characteristics. This is clearly not an attractive assumption. The product characteristics that we are able to observe are clearly not set at random. It is natural to conjecture that the characteristics that we do not see are also chosen strategically. No natural IV strategies are available for this problem.

In the paper, much of the emphasis is placed on allowing for consumer heterogeneity. In many applications, much of the variation that we see in prices is due to firm heterogeneity. Allowing for a more flexible specification that estimates firm level production functions is desirable. However, finding appropriate identifying assumptions for the supply side can be difficult as well.

Summary

Yang, Chen and Allenby have produced a fine piece of work. Putting standard models of supply and demand in differentiated products markets into a Bayesian framework is a useful contribution. Their framework is attractive numerically, facilitates decision making and allows us to incorporate a priori information into our estimates. However, there are still limitations to this framework, many of which are shared by other models in the literature. I look forward to future applications where Bayesian approaches are used to estimate the primitive supply and demand parameters in these markets.

References

Albert, J. and S. Chib. (1993). ''Bayesian Analysis of Binary and Polychotomous Response Data'', Journal of the American Statistical Association 88, 669–679.

Aguirrebagiria, V. and P. Mira. (2003). ''Sequential Simulation Based Estimation of Dynamic Discrete Games'', unpublished working paper, Boston University.

Bajari, P. and A. Hortacsu. (2003). ''The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions'', Rand Journal of Economics pp. 329–355.

Bajari, P. and L. Ye. (2003). ''Deciding Between Competition and Collusion'', forthcoming in The Review of Economics and Statistics.

Berry, S., J. Levinsohn, and Pakes, A. (1995). ''Automobile Prices in Market Equilibrium'', Econometrica 63(4), pp. 841–890.

Berry, S., M., Ostrovsky, and A. Pakes. (2003). ''Simple Estimators for the Parameters of Dynamic Discrete Games (with Entry/Exit Examples)'', unpublished Harvard University working paper.

Chib, S. and E. Greenberg. (1995). ''Understanding the Metropolis-Hastings Algorithm'', American Statistician 49, 327–335.

Davis, P. ''Spatial Competition in Retail Markets: Movie Theaters'', unpublished working paper, London School of Economcis.

Gelman, A., D. Rubin, and H. Stern. (2003). Bayesian Data Analysis, Second Edition CRC Press.

Geweke, J. (1997). ''Posterior Simulators in Econometrics'', in D. Kreps and K.F. Wallis (eds.), Advances in Economics and Econometrics: Theory and Applications, III Cambridge: Cambridge University Press, 128–165.

Geweke, J., M. Keane, and D. Runkle. (1994). ''Alternaitve Computational Approaches to Inference in the Multinomial Probit Model'', Review of Economics and Statistics 76, 609–632.

Geweke, J., G., Gowrisankaran, and R. Town. (2003). ''Bayesian Inference for Hospital Quality in a Selection Model'', Econometrica 71, 1215–1238.

Leslie, P. (2003). ''Price Discrimination in Broadway Theatre'', forthcoming in the Rand Journal of Economics.

McCulloch, R. and P. Rossi. (1994). ''An Exact Likelihood Analysis of the Multinomial Probit Model'', Journal of Econometrics pp. 207–240.

Pesendorfer, M. and P. Schmidt-Dengler. (2003). ''Identification and Estimation of Dynamic Games'', NBER Working Paper W9726.

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

Petrin, A. (2002). ''Quantifying the Benefits of New Products: The Case of the Minivan'', Journal of Political Economy 110, 705–729.

Rossi, P., R. McCulloch, and G. Allenby. (1996). ''On the Value of Household Information in Target Marketing'', Marketing Science 15, 321–340.

Tamer, E. (2003). ''Incomplete Simultaneous Discrete Response Model with Multiple Equilibria'', Review of Economic Studies 70(1), pp. 147–167.

Villas-Boas, S. (2003). ''Vertical Contracts Between Manufacturers and Retailers: An Empirical Analysis'', unpublished working paper, University of California at Berkeley.

Yang, S., Y. Chen, and G. Allenby. (2003). ''Bayesian Analysis of Simultaneous Supply and Demand'', this issue, Quantitative Marketing and Economics.

Quantitative Marketing and Economics, 1, 285–291, 2003.

Comment

STEVEN BERRY
Deptartment of Economics, Yale University and NBER

1. Introduction and background

Yang, Chen and Allenby propose a Bayesian method for estimating a class of equilibrium market models that are similar to the ‘‘BLP’’ model of Berry, Levinsohn and Pakes (1995). Building on earlier work in the Industrial Organization and discrete-choice econometrics literature, BLP proposed a model of differentiated products markets with discrete choice demand, unobserved product characteristics on the demand and cost side and multi-product oligopoly pricing. Issues of methodological importance included the econometric endogeneity of prices and the role of observed and unobserved consumer heterogeneity.[1] A number of more recent papers have extended the framework in various ways, including modifications to the model so that it is more appropriate to the questions of wholesale and retail pricing that are common in the marketing literature (Sudhir, 2001; Villas-Boas, 2002).

In this class of models, a Bayesian approach has a number of important potential benefits, especially in the context of work on products and markets, where the number of observations is clearly finite.

In this comment, I will very briefly review the earlier ‘‘classical’’ econometric work, turn to a discussion of the advantages and trade-offs of Bayesian methods in this class of models and then discuss some particulars of the Yang, Chen and Allenby approach, both in methodology and empirical application. I have a number of questions about the approach, including what assumptions are necessary to separately identify the demand unobservables as parameters. I conclude that, despite these open questions, the paper makes a very valuable contribution in presenting a coherent Bayesian approach to this class of models.

1.1. *Classical approaches*

In most work on BLP style models, the demand and cost unobservables are assumed to be uncorrelated with a set of exogenous instruments (see BLP) and/or are assumed to be restricted by a set of panel-data style restrictions, as in Nevo (2001). Estimation

1 Other possibly important issues, like dynamics and imperfect information, are typically not treated in this literature.

then typically proceeds by the Generalized Method of Moments (GMM) of Hansen and Singleton (1982). The "instruments" used include measures of changes in choice sets and measures of input prices. Panel data restrictions could be across markets (as in Nevo, 2001) or across time within products.

While most existing work is in the GMM framework, Berry, Levinsohn and Pakes (2001) note that with consumer-level data, one could estimate the demand model by Maximum Likelihood (ML), treating the demand unobservables as parameters to be estimated from the multiple observations on consumer choice within a market. However, they warn that these unobservables are typically perfectly co-linear with the "mean" effect of product characteristics, including price. Thus, a further product-level analysis is necessary to untangle these mean effects from the effect of the unobservables; the problems of simultaneity and "instruments" reappear in this second-stage analysis.

The discussions of estimation, model fit and "identification" in these models have highlighted several issues: the importance of unobservable consumer heterogeneity, the importance of product-level unobservables (and the associated problems of price endogeneity) and the importance of the choice of instrumental variables and/or panel data restrictions on those unobservables. Intuitively, in order to estimate demand price effects and substitution patterns, there has to be some exogenous source of change in prices and in the choice set facing consumers. The latter is needed to estimate realistic demand substitution patterns. It is important to note that these basic issues do not go away when moving to a Bayesian method, at least if we do not want the prior and the functional form restrictions to have an undue affect on the resulting posterior.

2. Advantages and trade-offs of Bayesian methods for equilibrium discrete choice models

Before turning to a slightly more detailed discussion of the Yang, Chen and Allenby paper, I will outline some advantages and disadvantages of Bayesian methods in the context of BLP-style equilibrium models.

2.1. Advantages

As compared to GMM and ML methods, Bayesian methods offer several, possibly important, advantages when applied to this class of models.

First, they provide a estimation method that does not rely on large sample theory. This is potentially important because a typical differentiated products dataset has only a moderate number of products and markets (perhaps in the tens or hundreds, but very rarely many thousands). The stronger assumptions of Bayesian analysis may yield more "precise" estimates (posteriors) than classical methods, an advantage when datasets are not overwhelmingly large. A number of authors in the current literature report some problems in obtaining precise estimates of, say, the variance of tastes for a particular product characteristic (as in the original BLP.) This is particularly true in studies that use product and market-level "aggregate" datasets. Even when GMM methods seem to work well, the asymptotic approximations to the standard errors are sometimes misleading. Thus, compared to, say, labor economics or demography, Bayesian methods may be more important when studying data on products and markets.

Second, recent Bayesian methods may offer computational advantages. With a large number of products, existing classical methods can require a large number of simulation draws to estimate the market shares with sufficient precision (Berry, Linton and Pakes, 2003). Some authors in the broader discrete-choice literature report a lower computational burden using recent advances in Bayesian computation, such as Monte Carlo chain methods. Furthermore, classical methods require one to search over the parameter space to minimize the objection function, which can be difficult and time-consuming in such non-linear models.

Third, in policy analysis it may be convenient to consider the role both of prior beliefs and of policy-makers' loss-functions. These are natural in the Bayesian context. One could consider very interesting applications of Bayesian methods to, say, merger analysis.

2.2. *Trade-offs*

As is typical, these advantages of Bayesian estimation routines come at the cost of stronger assumptions. Fully parametric distributional assumptions on unobservables replace the zero-covariance moment restrictions of GMM. The current literature already comes under criticism for relying on functional form assumptions and Bayesian methods require a yet stronger set of parametric assumptions.

With a Bayesian method, it now becomes necessary, rather than optional, to have a correctly specified supply-side. An advantage of obtaining demand estimates without supply-side restrictions is that we often lack confidence in any particular supply-side model of the interactions between oligopolists and between retailers and wholesalers.[2] Even though supply-side estimates are required to answer most policy questions, it might help if at least the demand-side parameter estimates are unaffected by supply-side misspecification. Further, GMM supply side estimates are typically very easy to obtain conditional on demand-side parameters, so a robustness analysis of various supply-side assumptions is typically very easy and quick.

2 There are a few questions of interest that do not require supply-side estimates. For example, Berry, Levinsohn and Pakes (2001) consider the question of new product introduction (or withdrawal) conditional on current prices. In economics, the computation of ideal utility-based price indexes is another question that does not require knowledge of supply-side estimates.

In realistic market settings, Maximum Likelihood (ML) and Bayesian methods (but not GMM) face the problem of complicated non-linear implicit first-order conditions that define prices. In this case, the distribution of prices must be found via a non-linear change-of-variables from the distribution of the unobservables. When the densities of some variables are only defined via an implicit change-of-variables argument, taking random draws from those distributions can be time-consuming. This difficulty may partly or totally offset the Bayesian computational advantages that are reported in simpler settings. Summarizing the computational issues, then, it is not clear whether one method has an advantage over the other; further research might be needed here.

More subtly, ML and Bayesian methods require that there is a unique equilibrium in the data, as noted, for example, by Amemiya (1985). If there is the possibility that a given set of exogenous observable and unobservable variables could be associated with a different equilibrium set of prices and quantities, then there is no longer a one-to-one map between the unobservables and the endogenous prices (conditional on the exogenous observables and the demand errors) and so the change-of-variables necessary to define the likelihood is no longer correct.

Note that in the case of multiple equilibria, there is effectively a missing variable that selects the observed equilibrium from the various possibilities. This is not a problem for GMM, which requires only that we can compute the unobservables as a function of the parameters, but it is a problem for any method that requires a likelihood. Thus, one additional assumption for Bayesian analysis is that either [i] there is a unique equilibrium to the supply relationship or [ii] there is an equilibrium selection mechanism that always selects a unique equilibrium from the possible set, as a function of the exogenous (observed and unobserved) variables.

3. Implementation in Yang, Chen and Allenby

I will make a few comments about the particular likelihood and supply-side first-order conditions used in Yang, Chen and Allenby and then discuss the empirical example.

3.1. *The likelihood and the first-order condition*

There seem to be two important points in deriving the likelihood. The first point is to treat the demand unobservables (at both the consumer and product level) as latent parameters; the authors derive a posterior for these parameters rather than treating them as simply stochastic components of the model with a known distribution. (The cost errors are treated as traditional random shocks.) The second point is to recognize that the first-order conditions define an implicit reduced-form for price (assuming uniqueness of equilibrium).

The first decision allows Yang, Chen and Allenby to condition on the unobserved demand factors, ξ, in the likelihood. This is similar to the first step in the ML or GMM approach of Berry, Levinsohn and Pakes (2001) and very similar to the traditional discrete choice literature that models "alternative-specific constants" as parameters to be estimated.

Berry, Levinsohn and Pakes (2001) discuss the identification problem that arises because the unobserved ξ_j are perfectly co-linear with the "mean" effect of characteristics x and price. That is, in equation (1), we could change the means of the β_i's and α_i and then make a perfectly offsetting change in each ξ_j, leaving every implication for demand unchanged. Berry, Levinsohn and Pakes (2001) propose a second-stage analysis where further restrictions are placed on the estimated "mean utilities" so that the separate effect of price and x can be estimated; these restrictions could take the form of instrumental variables restrictions, panel data restrictions or prior information.

In a Bayesian context, the identification problem is somehow "solved" by the prior and the various functional form restrictions, but it would be nice to see some discussion of how the Yang, Chen and Allenby framework avoids the fundamental problem. I should emphasize that this is not a problem with Bayesian analysis in general, but rather with any approach that treats the ξ as parameters. Bayesian approaches can probably make use of the same sort of restrictions that work elsewhere and can probably make even better use of the "prior" information restrictions discussed by Berry, Levinsohn and Pakes (2001).

The Yang, Chen and Allenby likelihood conditions on ξ because it is a "parameter" for each product, but that still leaves the problem of endogenous prices. The solution is to derive the implicit density for prices using the change-of-variables rule on the map between the unobserved cost shocks and the prices. Given a unique equilibrium, the first-order conditions are a unique map conditional on exogenous observed data and on ξ. This is a straightforward solution to a difficult problem.

Using repeated observations on the same household to estimate the unobserved household tastes is similar to the ML approach of Goettler and Shachar (2001) and earlier authors, but the Bayesian approach is probably computationally much easier. Note that household-specific taste parameters may be difficult to estimate without either many repeated choices per household or else prior information.

Yang, Chen and Allenby assume that supermarkets have access to similar techniques: their retailers are assumed to set optimal prices conditional on the unobserved household tastes. This very strong assumption, which seems necessary to avoid a difficult integration problem on the supply side, could use more discussion.

Aside from the concerns about identification, the approach to the likelihood is quite nice. It should be noted that this likelihood is as useful for ML analysis as for Bayesian analysis. There is some talk in the paper of the necessity, in an ML approach, of integrating out the unobservables in supply and demand, but the authors own likelihood shows that this is in fact not necessary. Again, an ML approach might have difficultly optimizing with respect to all those parameters, particularly the household specific tastes.

On the supply side, the authors sometimes talk as though the use of a linear reduced-form for the pricing equation (the "limited information" model) somehow avoids the problems of fully specifying the supply-side. There is an implied analogy to single-equation instrumental variables methods. Of course, in an ML or Bayesian context, the linear supply relationship (as a function of "instruments") is only appropriate if the true supply relationship reduces to that linear equation, a very unlikely event. Thus the "instruments" could be correct while the method is still wrong for functional form reasons. Treating the "limited information" model as somehow avoiding the problem of supply-side specification seems misleading to me. It might be more appropriate the treat the linear-in-instruments model as a kind of robustness check.

3.2. The empirical example

The empirical example is more in the nature of a proof-of-concept of the method as opposed to a real empirical study, which strikes me as just fine in a methodology paper, as long as the results are treated with appropriate care. There are only three products per market, which contrasts greatly with many earlier non-Bayesian papers that handle a much larger number of products. In the full model, the demand and supply errors are assumed to be uncorrelated with each other (which is highly restrictive) and there is no treatment of correlation in the product-level errors over time. The "instruments" in the linear supply model are implausible in the face of such correlation across time.

In fact, the model is so very simplified that one wonders about the actual computational burden of the model—if the model is easy to compute, then why employ such a simplified example?

The authors find, in common with almost every similar study, that consumer heterogeneity is important. Another conclusion of the study is that prices are indeed endogenous in the sense of being positively correlated with the unobservables. However, they find that the estimated demand elasticities are not much different when they assume the endogeneity away. There is little reason to believe that this last result will generalize across markets; other authors find very different results in different markets, using weaker assumptions. Further, the level of the demand elasticities is exactly the point where the identification question of MicroBLPCowles comes into play.

It would be nice to see some discussion of how to estimate the model on aggregate (market level) data; it seems on the surface that integration with respect to household unobservables would again play a role.

4. Conclusion

This is a valuable paper that, to my knowledge for the first time, presents a viable Bayesian method for estimating "BLP" style models of equilibrium discrete-choice differentiated products markets. Bayesian methods have important possible advantages for the study of product markets generally, especially because the number of products and markets is typically not overwhelmingly large and so the small sample advantages of the Bayesian approach may be important.

Treating the demand unobservables as parameters is one important decision made by the authors. The implications of this decision for identification could use further discussion. Other issues that could use more discussion include how to implement the model on market-level data and in models without perfect price discrimination by household type.

It would be nice in future work to see comparisons between Bayesian and non-Bayesian methods in terms of estimates, precision, computational burden and sensitivity to assumptions about functional form and the distributions of unobservables. This work would usefully involve both Monte Carlo studies on artificial data and the use of a variety of real-world datasets. We may find that different methods have advantages on datasets of different sizes and types or that the attractiveness of the methods varies with the willingness of the researcher to impose different sets of assumptions.

References

Amemiya, T. (1985). Advanced Econometrics. Harvard University Press.

Berry, S., J. Levinsohn, and A. Pakes. (1995). "Automobile Prices in Market Equilibrium", Econometrica 60(4), 889–917.

Berry, S., J. Levinsohn, and A. Pakes. (2001). "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market", working paper 1337, Cowles Foundation, forthcoming JPE, Feb. 2004.

Berry, S., O. Linton, and A. Pakes. (2003). "Limit Theorems for Differentiated Product Demand Systems", working paper 1372, Cowles Foundation, forthcoming Review of Economic Studies.

Goettler, R. L., and R. Shachar. (2001). "Spatial Competition in the Network Television Industry", RAND Journal of Economics 32(4), 624–656.

Hansen, L., and K. Singleton. (1982). "Generalized Instrumental Variables Estimation of Method of Moments Estimators", Econometrica 50, 1269–1286.

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

Sudhir, K. (2001). "Structural Analysis of Competitive Pricing in the Presence of a Strategic Retailer", Marketing Science 20(3), 244–264.

Villas-Boas, S.B. (2002). "Vertical Contracts Between Manufacturers and Retailers: An Empirical Analysis", Discussion paper, University of California, Berkeley.

Quantitative Marketing and Economics, 1, 293–298, 2003.

Comment

JEAN-PIERRE DUBÉ AND PRADEEP K. CHINTAGUNTA
Graduate School of Business, University of Chicago

The authors develop a Bayesian method to recover the structural parameters from an industry characterized by a heterogeneous logit demand system, on the consumer side, and the Bertrand-Nash equilibrium concept, on the supply-side. The logit demand system is augmented by including error components accounting for unobserved consumer heterogeneity, θ_i, as well as error components accounting for unobserved (to the researcher) product characteristics, ξ_t. By jointly modeling supply and demand, they devise a full-information approach that resolves the potential endogeneity bias in approaches that treat prices as exogenous. Calibrating their model using a household-level scanner panel data set, they draw two empirical conclusions. First, they find that observed prices are correlated with demand shocks, consistent with previous research. Interestingly, this correlation does not appear to lead to biases in the mean price sensitivity across consumers once unobserved heterogeneity is accounted for. This is in variance with previous research that has found such a bias. Second, the results suggest that the additional structure imposed by the supply-side improves the in-sample fit of the model.

Methodologically, the authors make an important contribution to the literature. The Bayesian approach resolves a difficult computational problem that arises in classical likelihood-based approaches. The classical approach assumes the random effects are drawn from a known parametric distribution. Estimation is carried out using the unconditional likelihood, which is integrated over the distribution of the random effects. Integrating over both the random effects as well as unobserved product characteristics ("demand shocks") complicates the evaluation of the unconditional likelihood function as the former vary only across consumers whereas the latter vary over time. The Bayesian approach obviates the need to evaluate this unconditional likelihood thereby making the problem amenable to empirical estimation. This methodological contribution should appeal to researchers modeling the supply-side to control for the endogeneity of marketing variables, and/or to conduct policy experiments using the estimated model parameters. Use of a likelihood-based method also enables researchers to conduct structural tests on the supply-side.

The practical discussion of using full information versus limited information approaches to parameter estimation boils down to trade-offs. A balanced discussion of the relative strengths of alternative approaches to this same empirical problem that rely on fewer modeling assumptions will be a useful addition.

The empirical implementation of the proposed methodology also raises a few issues. Some of these can be resolved through more detail. It is unclear however, whether the data are suitable to showcase the proposed methodology appropriately. Overall, despite the strength of the methodological contribution, it is unclear whether we can make strong conclusions based on the empirical findings.

The remainder of our discussion will focus on two main areas.

1. The trade-offs involved in using full versus limited information approaches as well as maximum likelihood versus instrumental variable (IV) approaches.
2. Empirical implementation.

1. Trade-offs between full-information versus limited information and IV procedures

The authors position existing limited information and IV approaches as a compromise to avoid the technical problems discussed above. In practice however, these IV approaches are typically adopted as they generate consistent demand estimates under far more parsimonious model assumptions.

The full information approach requires taking a stand on the exact model generating the observed prices. In the paper, several static pricing models are compared. This additional structure constrains the estimated demand parameters, as they need to satisfy both the supply and demand equations. In principle, this additional structure leads to more efficient parameter estimates. However, model mis-specification on the supply-side can also lead to biased estimates. In contrast, a limited information approach (Villas-Boas and Winer 1999) remains agnostic about the underlying model generating observed prices. While the demand estimates are estimated with less precision, an inappropriate supply-side model does not contaminate them. Similarly, IV procedures (Chintagunta, Dubé and Goh 2003) do not require assumptions about the precise form of pricing. For policy analysis, one can estimate the supply-side parameters in a second stage that conditions on the demand estimates (Nevo 2000, Berto Villas-Boas 2003). In this way, the demand estimates are not contaminated by an inappropriate supply-side model.

The paper's view of IV approaches is that they are equivalent to imposing a supply-side model in which prices are a linear function of instruments. This point is well taken. However, to the extent that one can more readily identify exogenous factors that shift prices than one can describe the true nature of the price setting behavior of manufacturers and retailers, our ability to obtain unbiased estimates is not compromised. Further, one can use a flexible functional form instead of a linear regression. Indeed, developing a structural model from which one derives a pricing policy predicting occasional strategically-timed price cuts—a pattern observed in most retail scanner data—is extremely difficult. Existing theories that predict similar pricing involve models with a time-varying price elasticity of demand, on the consumer side, and a dynamic pricing game, on the supply-side. Solving the dynamic programming problem of retailers is beyond the scope of the current paper. However, it is not clear whether the static models considered provide a satisfactory explanation of the observed price variation over time in the data. The key question that one needs to answer is—are we really worse off with less efficient estimates that do not impose inappropriate assumptions on the process generating prices?

2. Empirical implementation

In the empirical section, the authors use scanner data to calibrate the model and demonstrate the proposed methodology. Several surprising results are documented. First, the authors find evidence of correlation between the unobserved product attributes, ξ_{jt}, and prices; however they do not find any evidence of endogeneity bias. Second, they find support for a pricing model that is based on an information-set different from that assumed in the literature. Third, the paper finds substantial improvement in model fit in moving from the limited to the full information approach when heterogeneity is accounted for. In the following, we first discuss the data used in the paper and then the above findings.

A concern that arises is the unusually small data set used to capture features of both supply and demand in this industry. For instance, with 92 weeks, 631 purchase occasions and 3 brands, there are only 2.3 observed choices for a given brand in a given week (on average). The authors estimate weekly brand-specific parameters, ξ_{jt}, to control for unobserved product characteristics. The estimation of these parameters relies either on the information in the data (the 2.3 observed choices for a given brand in a given week) or the information contained in the prior. Given how few choices we have per brand in a typical week, it would help to have a discussion on how well we can inform ourselves of the distribution of ξ_{jt} with the data at hand. Since classical estimation of these parameters would be hopeless with so few observations, the Bayesian approach is indeed better suited to handle this type of small sample problem. But, it seems crucial at this point to discuss the sensitivity of the prior chosen. It would have been helpful to report whether the posterior distribution of ξ_{jt} deviated far from the prior. Similarly, it would have been helpful to see how sensitive the posterior distribution would be to a different choice of prior. Given how little information the data contain about specific realizations of ξ_{jt}, it is unclear how much we can learn about the aggregate correlation between prices and ξ_{jt}. Only in weeks with unusually low prices would one expect to observe enough choices of a brand to learn about that brand's unobserved attribute. In fact, referring to the scatter plots at the end of the paper, we see that the documented correlation between prices and ξ_{jt} is most pronounced at the lowest price levels.

A surprising empirical finding is the stated lack of endogeneity bias. This result obtains despite the positive correlation between prices and ξ_{jt} (between 0.164 and 0.39), and the preponderance of evidence to the contrary in related work. Typically, research on this topic has focused on biases in the mean price sensitivity. The authors find that controlling for the supply side does not impact the estimated mean price sensitivity. Does this mean the endogeneity bias is unimportant? In the case of a linear regression, it is straightforward to characterize the endogeneity bias. Characterizing the bias is not straightforward in the context of the non-linear model used. First, it is unclear how strong the correlation between prices and ξ_{jt} must be to generate statistical bias. Perhaps another useful statistic to report would be the marginal impact of the estimated ξ_{jt} on demand. It is also unclear how the endogeneity bias will manifest itself in the estimates. For instance, Chintagunta et al. (2003) also document that controlling for price endogeneity leads to different results for heterogeneity. The authors report two moments of the parameter distribution; but they do not compare the entire posterior distribution of tastes. It might have been helpful to look at a histogram of the posterior distribution of parameters to see if it is comparable. Perhaps a simulation exercise would have been more instructive than the calibration to scanner data to try and characterize the estimation bias. Also, the authors could have compared the marginal effect of prices on choices under the 10 models (e.g. own price elasticity) as a more informative metric to assess the effect of endogeneity.

Another unusual finding in the paper is the support for a model in which firms are assumed to aggregate demand over the posterior distribution of heterogeneity based on the sample. Typically, researchers integrate over an assumed population distribution of heterogeneity (e.g. the normal distribution using the hyperparameters for the mean and variances).[1] The authors' finding provides the basis for an interesting discussion about the information sets available to firms. While it may be too strong to assume the population distribution of tastes is known, at the same time, do we truly believe that retailers and manufacturers set prices based on this exact panel data set? In general, the finding opens a potentially interesting discussion about the information used by manufacturers and retailers when they set prices. A more thorough investigation of the appropriate information sets facing firms seems like an interesting area for future research.

The authors also observe that the limited and full information approaches result in different price effects in the no heterogeneity case whereas the two models result in similar estimates when heterogeneity is accounted for. They conclude, "mis-specification of heterogeneity can mask the importance of the supply-side model." However, after accounting for heterogeneity, the estimates from the full information model become close to those from the "no endogeneity" specification, whose results are similar to the limited information case under both heterogeneity and no heterogeneity specifications. A plausible re-interpretation of the results would be that accounting for heterogeneity mitigates the bias resulting from imposing an incorrect full information supply-side model. For instance, the IIA assumption in the homogeneous logit demand system restricts a brand's price-cost margin on the supply-side to be proportional to its market share. However, in the data (Table 1),

1 The results from this second model are not reported. However, the authors state that it provided an inferior fit to the data.

Coors Light has the largest share on average; but not the highest price.[2] The inclusion of heterogeneity offsets the IIA property and its consequent effect on the pricing model.

A related finding is the superior fit of the same full-information model versus a limited information approach (model M9 versus M10), although the fit of the limited information approach is superior when consumer heterogeneity is not modeled. The intuition for this improved fit is unclear. Looking at the results in Table 2, we see that accounting for heterogeneity in the limited information model improves the log-marginal density from −3400 to −2464. Since accounting for heterogeneity does not influence the supply-side in this model, all the improvement in fit can be attributed to the demand model. In the full information case, where accounting for heterogeneity does influence the pricing model, adding heterogeneity improves the log-marginal density from −3410 to −2377 which is more than the improvement in the limited information case. This suggests that heterogeneity is also helping the fit of the pricing model in the full information case. As noted previously, it does appear that accounting for heterogeneity resolves the potential mis-specification of the supply-side model in the full information case.

Finally, lagged price is used as an instrument for price in the limited information specification of the model. As noted previously, the no endogeneity finding, comparing the results from the limited information model with those from the specification that ignores endogeneity, is conditional on the chosen instruments for price. One concern is whether these instruments are appropriate. A discussion of this issue, in the light of other types of instruments chosen by researchers (factor prices, wholesale prices, prices in other markets, etc.), is warranted. It would also be helpful to have more information about the quality of the instruments, such as an R^2. To assess the validity of the instruments, the authors could check the correlation between the estimated ξ_{jt} from the full information model (M9) with the chosen instruments, i.e., lagged prices.

In summary, we reiterate that we believe that this paper makes an important methodological contribution to the burgeoning area of understanding market demand and supply behavior. We look forward to further research in this area that can shed some light on the issues raised in this discussion.

References

Berto Villas-Boas, S. (2003). "Vertical Contracts Between Manufacturers and Retailers: An Empirical Analysis", Discussion paper, University of California—Berkeley.

Chintagunta, P.K., J.-P. Dubé, and K.-Y. Goh. (2003). "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-level Brand Choice Models", working paper, University of Chicago.

2 We compare prices as the estimates marginal costs (brand intercepts) are all statistically the same (M3, M4, M8 and M9).

Nevo, A. (2000). "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry", The RAND Journal of Economics 31(3), 395–421.
Villas-Boas, J.M. and R. Winer. (1999). "Endogeneity in Brand Choice Models", Management Science 45, 10, 1324–1338.

Quantitative Marketing and Economics, 1, 299–304, 2003.

Reply to Comments on "Bayesian Analysis of Simultaneous Demand and Supply"

SHA YANG
Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012
E-mail: shayang@stern.nyu.edu

YUXIN CHEN
Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012
E-mail: ychen@stern.nyu.edu

GREG M. ALLENBY
Fisher College of Business, Ohio State University, 2100 Neil Avenue, Columbus, OH 43210–1144
E-mail: allenby.1@osu.edu

We thank the reviewers for a provocative and interesting set of comments. The comments touch on a number of philosophical differences in analysis that has stimulated our thinking. By bringing these assumptions to our attention, the reviewers have done us a service in deepening our understanding of important issues. We hope that this exchange of perspectives provides readers with a deeper understanding of the issues they face in the analysis of data.

Response to Steven Berry

Professor Berry identifies a number of tradeoffs encountered in using Bayesian methods for analyzing equilibrium discrete choice models, including the need to employ stronger parametric assumptions, the need to specify the supply-side model, and the assumption of a unique equilibrium in the data. Before addressing the specific points he raises, it is first useful to review some of the implicit assumptions behind likelihood-based inference, including Bayesian analysis.

The likelihood is a statement about the process assumed to generate the data. In models of economic behavior, the likelihood reflects a process of constrained utility maximization on the part of the consumer, and profit maximization on the part of the supplier. The likelihood is a device that links stochastic innovations in the model specification (e.g., error terms) to the observed data. That is, the distribution of the observed data is derived from primitive assumptions the error term. In a random utility model, the error term results in uncertainty in the slope of the indifference curves and the resulting utility maximizing choice.

Our view is that likelihood-based inference offers a distinct advantage over GMM because it makes precise statements about the probability of the observed data. In a likelihood-based analysis, the analyst is confronted with the correspondence between the model and data, and cannot fit a model that is not supported by the data. For example, the assumption of a Cobb-Douglas utility function in the analysis of discrete choice data would be problematic because this utility function does not lead to corner solutions and the likelihood is zero at all parameter values. However, the use of GMM estimator would yield estimates in this case because parameters do exist that conform to the zero moment conditions. While the GMM estimator requires less of the analyst and the data, it also delivers less precision. Our view is that precision will become increasingly important as we attempt to understand the origins of utility and its relationship to consumer behavior.

We understand the reluctance expressed by Professor Berry in specifying the supply-side model, particularly for setting prices of packaged goods in a grocery store. However, issues of endogeneity and the presence of a supply side models are likely to be important in marketing because decision variables (the 4 P's) are typically set with the objective of maximizing return on investment. An advantage of Bayesian analysis is that it provides a measure of model fit to assess the posterior odds of alternative supply-side models, even if these models are non-nested. GMM and classical (frequentist) statistical methods do not provide a measure of model fit that can be translated into the probability that a model is correct.

Professor Berry raises an interesting point about the assumption of a unique equilibrium. This assumption is needed to derive the distribution of observed prices. If a unique equilibrium did not exist, then it would not be possible to derive the Jacobian for the change-of-variable calculus. GMM does not require the existence of a unique equilibrium—draws of the error terms can be simulated and used to derive the expected value of price even if multiple equilibrium exist. However, even though GMM can estimate model parameters in the face of multiple equilibrium, these estimates cannot be used to derive policy implications and predictions without knowledge of which equilibrium exists. GMM therefore does not completely solve the unique equilibrium issue. Moreover, Bayesian methods can always be applied to a limited-information reduced form supply-side specification where multiple equilibrium do not exist (e.g., model 10). We hope that future research will shed greater light on the multiple equilibrium issue, including the frequency it occurs and its affect on parameter estimates and policy implications.

Our Bayesian analysis exploits the method of data augmentation in which realizations of the demand shocks are treated as latent parameters. This approach is similar to Berry, Levinsohn and Pakes (2001) who estimate the demand shocks (ξ_j) in the first stage of a two-stage analysis. The demand shocks are identified in our analysis because they have a mean of zero and unrestricted covariance matrix (equation (4)). The assumption of a zero mean does not allow tradeoff with the mean of random-effect distribution. In BLP (2001), the demand shocks are estimated as period and brand specific constants, and the zero-mean assumption is not assumed part of the model structure during the first stage.

Data augmentation is also used to estimate the unobserved household taste parameters, θ_i. As discussed in the appendix, the estimation algorithm proceeds by recursively generating draws from the full conditional distribution of all model parameters, including the augmented taste parameters. Theoretically, the posterior distribution of a specific household's taste parameters, θ_k, is equal to the integral of equation (28) with respect to all model parameters except θ_k. A remarkable aspect of MCMC estimation is that the marginal posterior distribution for θ_k is obtained as a by-product of the estimation procedure—the analyst simply has to save the draws of any model parameter, and conduct inference through the simulated realizations by computing means, variances and confidence intervals using the simulated values.

We acknowledge that a strong supply-side assumption is made regarding the setting of prices based on unobserved household tastes. A more standard assumption is that retailers and manufacturers set prices based on the distribution of tastes ($\bar{\theta}$ *and* $\sum_\theta$), not the individual realizations θ_i. As discussed in the paper, the stronger assumption leads to a better model fit. It also leads to a Markov chain with less auto-correlation and quicker convergence. We believe this improvement may be due to linking the supply and demand equations with parameters that are close to the data, rather than an additional step away from the data in terms of the model hierarchy. Additional research is required to understand this issue more completely.

Finally, with regard to the empirical evidence provided in the paper, we feel that there are many avenues for extension and improvement, including the specification of more general covariance matrices, state dependence, and the examination of better instruments. Despite the relative simplicity of the analysis (3 brands, 92 weeks of data, 185 households), the computation burden was fairly severe—model 9, the best fitting model, took 75 seconds per iteration of the chain on a Pentium 3.0 GHz machine using a standard C + + complier. However, this burden will lessen as computer speeds continue to increase, allowing analysts to examine additional datasets with more complex models of behavior, facilitating the documentation of empirical results across multiple datasets.

Response to J.P. Dube and Pradeep Chintagunta

Professors Dube and Chintagunta raise important issues regarding the tradeoffs of using a full information versus a limited information approach, and the results of our empirical study. If a goal of an analysis is to understand how resources are allocated by managers and firms, then it is necessary to model the supply side. The proposed Bayesian approach offers a viable approach to estimating models that will likely arise in the context of marketing applications. It also offers a mechanism for testing alternative supply-side specifications to determine effects on parameter estimates. If the analysis has the less ambitious goal of controlling for the possible effects of the allocation decision, then a limited information approach may appear to be preferred because it does not require a supply-side model.

The issue of whether the analyst is really worse off with less efficient estimates that imposes fewer assumptions is important. Assessing this tradeoff requires information on the loss of efficiency. Unfortunately, instrumental-variable approaches have unknown finite sample properties so this tradeoff cannot be assessed. While the instrumental-variable approach is consistent and asymptotically normally distributed, the point at which the asymptotic properties take hold is dependent on many factors, including the model structure, nature of the dependent variable, and the ability to identify suitable instruments. Little practical guidance exists for the selection of instruments and their effect on the finite sample properties. We believe that in the context of many marketing and economic applications, where many "units" of analysis but limited information per unit characterize panel data, the use of limited-information methods is not advised without simulation analysis to assess their finite-sample properties.

Our empirical results indicate that positive correlation between prices and the demand shocks (ξ_j) do not translate into biased estimates of price sensitivity. Instead, we find that alternative model specifications lead to changes in the model intercepts (see Table 3). This result is contrary to previously published studies that have identified the correlation as a source of the bias in the price coefficient. The demand shocks enter the model as an additive term in equation (1), and it seems reasonable to us that their omission would most affect the model intercepts. Past studies using likelihood-based inference have not been successful in incorporating unobserved heterogeneity, and this may be the reason for the conflicting results.

Questions are raised about the role of the prior distribution on posterior estimates. Individuals not working in this paradigm often raise this concern, and the presence of a prior distribution is seen by many as a weakness of the Bayesian approach. A proper prior distribution is needed to ensure the existence of the posterior, but is chosen in such a way that has minimal influence on posterior estimates. The influence of the prior is easy to check by conducting sensitivity analysis on the parameters of the prior distribution.

There are two distinct advantages to Bayesian estimation. First, Bayes theorem is used to account for all the uncertainty within the model structure, in contrast to the multi-step method of, for example, Berry, Levinsohn and Pakes (2001). Brand and time-specific constants can be estimated in the first step with much uncertainty, particularly in datasets like ours where there are so few observations. This uncertainty is not taken forward into the second step of analysis, rendering analysis with these methods prone to outliers, and the analyst is forced to develop intuition about the sequential behavior of the estimator (e.g., information about ξ_j is most pronounced at lowest prices). In contrast, hierarchical Bayes methods using Markov chain estimation automatically track this uncertainty, which is reflected by the spread of the posterior distribution of model parameters. It is the accurate accounting for uncertainty that facilitates small-sample analysis, rather than the extra information brought to the analysis through the prior distribution.

The estimated correlation between the demand shocks and price illustrate a second advantage of Bayesian methods. The model structure described by equations (22)–(27) does not impose any positive or negative correlation. The reported positive correlation is a result of the data, not the prior distribution (i.e., factor π_4 in equation (28)). We report similar results for the posterior estimates of household parameters θ_i that are not distributed normal, even though they are assumed normal in equation (24). The posterior distribution need not be normal because θ_i also appears in equation (25), and the posterior distribution is dependent on both equations. Bayesian methods offer flexibility in exploring departures of the data from prior assumptions imparted through the model structure and prior distributions, and do not determine the results except in limiting cases.

We report results for models 1–5 for completeness since likelihood-based studies have not been successful incorporating unobserved household heterogeneity. We view these models as mis-specified from practical considerations (all individuals are different) and an extensive set of published work documenting the importance of accounting for heterogeneity. With regard to Professor Dube and Chintagunta's reinterpretation of the results, we note that the proposed Bayesian approach allows one to test alternative supply-side models to determine which is best, and that we find the price coefficient to be similar with the limited-information specification. As with any result based on one dataset, additional analysis is needed to verify our empirical findings. Such analysis will hopefully explore a better set of instrumental variables. In our analysis, the use of lagged price resulted in an R^2 of 0.17.

Response to Patrick Bajari

Professor Bajari provides many useful perspectives on the use of Bayesian methods. We wish to add to the discussion of prior distributions by reminding readers that all analysis is subjective. Model specification involves many subjective decisions, as does the collection and reporting of data. Science is concerned about the ability to interpret, use and reproduce results, not whether the analysis is free from subjectivity. An advantage of Bayesian methods is that it facilitates incorporating subjective information into an analysis, allowing the analyst to explore the consequence of the many modeling assumptions. For example, sales are assumed to be a decreasing function of price in most economic demand models. Downward-sloping demand curves can be specified in an analysis using hard restrictions in the likelihood function (e.g., $y = \beta_0 - \exp(\beta_1) \times \text{price} + \varepsilon$), or with stochastic information expressed through the prior distribution (e.g., $\pi(\beta_1)$ is a distribution with mass centered in a negative region). Bayesian methods offer an additional tool for quantifying subjective beliefs so that analysis results can be reproduced.

The unresolved problems identified by Professor Bajari point to opportunities for future research. Problems and open issues remain in assessing convergence, specifying well-defined likelihoods, understanding the nature of endogeneity, and developing richer supply-side specifications. Allowing demand shocks to be driven by other variables, or possibly resulting from an optimal allocation of resources, can be modeled by adding additional layers to hierarchy. The issue of multiple price equilibrium is an excellent point and is discussed above in our response to Professor Berry.

Our analysis of household panel data is one of many possible applications of the Bayesian framework. While discussion has focused on the analysis of disaggregate panel data, Bayesian methods can also be used to analyze aggregated data where the distribution of heterogeneity is not directly inferred estimable. We hope our research, and the insightful comments by all three reviewers, will result in greater interest in jointly modeling resource allocation decisions by firms and consumer reactions in the marketplace, and other applications where endogeneity is present.