Journal of Marketing Research
Article Postprint
Volume XLV

A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data

Sungho Park
sp393@cornell.edu

Sachin Gupta
sg248@cornell.edu

Johnson Graduate School of Management
Cornell University
Ithaca NY 14853

May 2008

Sungho Park is a doctoral student and Sachin Gupta is Henrietta Johnson Louis Professor of Management and Professor of Marketing, both at the Johnson Graduate School of Management, Cornell University, Sage Hall, Ithaca NY 14853. We thank Vishal Singh and Yu Ma for sharing the data used in this paper, and to seminar participants at Cornell University for their helpful comments.

Abstract

A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data

We propose a Simulated Maximum Likelihood estimation method for the random coefficient logit model using aggregate data, accounting for heterogeneity and endogeneity. Our method allows for two sources of randomness in observed market shares – unobserved product characteristics and sampling error. Because of the latter, our method is suitable when sample sizes underlying the shares are finite. By contrast, the commonly used approach of Berry, Levinsohn and Pakes (1995) assumes that observed shares have no sampling error. Our method can be viewed as a generalization of Villas-Boas and Winer (1999) and is closely related to the "control function" approach of Petrin and Train (2004). We show that the proposed method provides unbiased and efficient estimates of demand parameters. We also obtain endogeneity test statistics as a by-product, including the direction of endogeneity bias. The model can be extended to incorporate Markov regime-switching dynamics in parameters and is open to other extensions based on Maximum Likelihood. The benefits of the proposed approach are achieved by assuming normality of the unobserved demand attributes, an assumption that imposes constraints on the types of pricing behaviors that are accommodated. However, we find in simulations that demand estimates are fairly robust to violations of these assumptions.

Key words: Random Coefficients, Logit Model, Endogeneity, Heterogeneity, Simulated Maximum Likelihood, Aggregate data, Brand Choice, Scanner Data.

Introduction

In the estimation of market demand, heterogeneity across consumers and the endogeneity of marketing activities, especially price, are major concerns of marketing researchers. It has been reported that ignoring heterogeneity and/or endogeneity causes a bias in demand estimates (Berry 1994; Keane 1997; Besanko, Gupta, and Jain 1998; Villas-Boas and Winer 1999; Chintagunta 2001; Chintagunta, Dubé and Goh 2005). Recognizing their importance, researchers have tackled both issues in aggregate models and disaggregate models.[1]

In disaggregate models, several estimation methods have been suggested and compared (Villas-Boas and Winer 1999; Petrin and Train 2004; Draganska, and Jain 2002; Yang, Chen, and Allenby 2003; Goolsbee and Petrin 2004; Chintagunta, Dubé, and Goh 2005). For aggregate models, Berry, Levinsohn, and Pakes (1995) developed a method (henceforth, BLP method) that provides consistent estimates under heterogeneity and endogeneity. The BLP method has been applied successfully in numerous studies (Sudhir 2001a; Petrin and Train 2004) and has become the most widely used approach for analyzing differentiated product markets.

A distinguishing feature of the BLP model is that it assumes that the observed market shares of alternatives have *no* sampling error. Randomness in shares in the BLP model is assumed to come *only* from unmeasured product characteristics. If the data being modeled contain more than minimal sampling error, the BLP estimator is not consistent and asymptotically normal (Petrin and Train 2004; Berry, Linton, and Pakes 2004). The BLP model was originally applied to automobile shares for the entire United States market in which the number of households was of the order of 100 million; hence, sampling error was negligible. In many subsequent applications of the model to weekly supermarket point-of-sale data (two examples are Chintagunta 2002 and Sriram et al. 2007), the underlying sample of shopper households is quite large, thus satisfying the no-sampling-error assumption of the BLP model.

In marketing one can identify a number of situations wherein the assumption of negligible or no sampling error in observed brand shares may not be tenable. These occur when the sample of shoppers underlying the observed shares is relatively small. Examples of such situations include the following:

a) Sales data from smaller retail stores;
b) Sales data for infrequently purchased categories;
c) Sales data at the stock keeping unit (SKU) level, which by definition have smaller sales than brands or brand-sizes;
d) Shares computed using aggregated household panel data. This may be necessary if the household-level data cannot be used due to, for instance, privacy concerns; and
e) Household panel data are aggregated to estimate brand shares because point-of-sale data are unavailable (e.g. Walmart does not provide point-of-sale data to ACNielsen or Information Resources Inc.).

In all these situations the assumptions of the BLP model may not met. We propose in this paper a Simulated Maximum Likelihood (SML) method to estimate an aggregate random coefficient logit model that considers endogeneity as well as heterogeneity. Our method is suitable for share data that are observed with sampling error. Thus, we assume that there are two sources of randomness in the model – unmeasured product characteristics, and sampling error. Our proposed method is motivated by the control function approach which was originally suggested for the disaggregate model by Villas-Boas and Winer (1999) and later extended to the aggregate model by Petrin and Train (2004). Villas-Boas and Winer (1999) developed their model for individual data and did not allow for unobserved individual heterogeneity. By contrast, our approach is for aggregate data, and we model unobserved heterogeneity using a random coefficients framework. In relation to Petrin and Train (2004), our model makes different assumptions on the distribution of the unobservables. We elaborate upon this distinction when we discuss the model in a subsequent section.

Using simulated data we demonstrate that the proposed estimator provides unbiased and efficient estimates of demand parameters. The estimation procedure is straightforward to understand and implement. Furthermore, the proposed method can readily incorporate other methods based on Maximum Likelihood Estimation (MLE). For example, we can incorporate Markov regime-switching models (or hidden Markov models) into our framework. By doing so, we can investigate parameter dynamics in choice models using aggregate data when both heterogeneity and endogeneity are present. A further benefit of our proposed model is that an endogeneity test statistic results as a by-product. A test for endogeneity based on the Wald statistic or the Likelihood Ratio statistic can then be easily performed.

In the proposed approach, we impose structure on the distribution of unmeasured product characteristics by making a normality assumption. We find that this leads to efficient estimates of heterogeneity parameters which are of great practical interest in marketing applications such as segmentation and targeting. The distributional assumptions we make impose restrictions on the types of pricing behaviors that are accommodated (we elaborate upon this later), although we find in simulations that demand estimates are fairly robust to violations of these assumptions.

Chintagunta, Dubé, and Goh (2005) showed that even when there is no price endogeneity, researchers have to pay attention to the presence of unmeasured product characteristics that affect consumer utility. Unmeasured product characteristics may include, for example, the impact of unobserved promotional activity, coupon availability, shelf space, national advertising, unquantifiable factors and systematic shocks to demand. If omitted from the model, the unmeasured product characteristics generate overstated variances in the estimated distribution of heterogeneity in household brand preferences and price sensitivities. An additional contribution of our paper is to expand upon this important finding of Chintagunta, Dubé, and Goh (2005). We show that problems due to the omission of the unmeasured product characteristics are more complex, and have additional facets which have not been reported in the literature. In particular, the omission can cause upward or downward biases in mean and/or heterogeneity parameters.

The remainder of the paper is organized as follows. In the next section, we review related literature. Following that, we present the model and explain our estimation method. We then evaluate the performance of the proposed method in simulation studies. In the next section, we apply the proposed method to scanner panel data and compare results with those from extant methods. We conclude in the last section.

Literature Review

We focus on literature that tackles endogeneity as well as heterogeneity in choice models.[2] In disaggregate models the available methods can be classified into three categories: 1) full-information maximum likelihood approaches (Sudhir 2001b; Draganska and Jain 2002; Yang, Chen, and Allenby 2003; Villas-Boas and Zhao 2005), 2) control function approaches (Villas-Boas and Winer 1999, Petrin and Train 2004), and 3) fixed-effect approaches (Goolsbee and Petrin 2004; Chintagunta, Dubé, and Goh 2005). In the full-information maximum likelihood approach, prices are modeled as the equilibrium outcome of a game between firms. By explicitly modeling price, one can integrate out unmeasured product characteristics and derive the unconditional joint likelihood of prices and choices. The control function approach is based on the concepts of Heckman (1978) and Hausman (1978), or it can be viewed as a reduced-form approximation of the equilibrium model. This approach requires two steps. First, the endogenous variable is regressed on instrumental variables. Second, the residual from the first step regression, or a function of the residual, is entered as an additional explanatory variable in utility to control for unmeasured product characteristics. In the fixed-effect approach, the first step is to capture the endogeneity by product- and/or market-specific fixed-effects and then, in the second stage, a standard instrumental variables method is applied to these fixed-effects. Chintagunta, Dubé, and Goh (2005) directly estimate the product- and market-specific fixed effects using MLE. Goolsbee and Petrin (2004) adopt the numerical inversion method (or contraction mapping) suggested by Berry (1994) to get the fixed effects.

There is a growing stream of work in marketing and economics that uses aggregate data to estimate choice models. One obvious reason for this trend is easier availability of aggregate data. A widely used approach for dealing with endogeneity as well as heterogeneity is the fixed-effect approach which was first developed for the aggregate model by BLP (1995) and later applied to a disaggregate model. Unlike the disaggregate model, in an aggregate model we cannot directly estimate the fixed-effects due to lack of degrees of freedom.[3] The BLP method circumvents the direct estimation of fixed-effects by using a numerical inversion method instead.

A weakness of the BLP method is its inability to recover heterogeneity parameters precisely when only aggregate data are used (Petrin 2002; Albuquerque and Bronnenberg 2006). Petrin (2002) proposed a technique to augment aggregate data with information relating consumer demographics to the characteristics of the products these consumers purchase. Similarly, Albuquerque and Bronnenberg (2006) supplement aggregate data with summaries of household switching behavior. An important strength of the BLP method is that it makes few assumptions about the distribution of unobserved product characteristics. The only assumption is that the unobserved characteristics are mean independent of the instrumental variables. As a result, the BLP method does not impose restrictions on the form of pricing behavior.

A number of recent papers perform Bayesian analysis of the random coefficient logit model using aggregate data. Musalem, Bradlow, and Raju (2007) consider two alternative scenarios that generate the observed aggregate data – one in which there are independent crosssections of consumers in each period, and the second in which there is a panel of consumers. They note computational limitations of their approach when the number of individual consumers underlying the aggregate data is larger than about 500. Their second scenario is similar to the one in Chen and Yang (2006) who propose a data augmentation approach to capture household heterogeneity. However, Chen and Yang do not consider unmeasured product characteristics or related price endogeneity issues, both of which are crucial to our research goals.

Recently, Jiang et al. (2007) propose a Bayesian analysis of the aggregate random coefficient logit model based on distributional assumptions about the unmeasured product characteristics. Similar to the BLP method, this approach is suitable when there is no sampling error in observed shares. Unlike our proposed SML method, model estimation in their approach requires inverting shares via the BLP contraction mapping as well as relatively complicated Markov-Chain Monte Carlo sampling. Like us, Jiang et al. demonstrate via numerical simulation that under misspecification of the distribution of the unmeasured product characteristic, their method continues to produce good results. In general, however, the properties of Bayesian estimators under model misspecification are not well established. By contrast, MLE is a strongly consistent estimator that minimizes the Kullback-Leibler Information Criterion (KLIC) (White 1982). That is, the proposed method provides estimates which are closest in KLIC to the true parameters in vector space defined by the normal approximation. In this respect we believe our method is more robust than Bayesian approaches.

Model and estimation procedure

Model

Our interest is in consistent and efficient estimation of the random coefficient brand choice model under assumptions of heterogeneity across consumers and endogeneity of marketing activities. We assume that consumers either choose a single unit of the brand that gives them the highest utility in the category or choose not to purchase in the category on a given shopping trip. In this paper, we focus on purchase incidence and brand choice behaviors only. In each week t=1,…,T, the utility of brand j=1,…,J for consumer h=1,…,H is given by the following expression:

$$u_{hjt} = x'_{jt}\beta_h + \xi_{jt} + \varepsilon_{hjt}, \tag{1}$$

$$u_{h(J+1)t} = \varepsilon_{h(J+1)t}, \text{ if no purchase,}$$

where x_{jt} is a k-dimensional vector of observed marketing mix variables and intrinsic brand values (brand intercepts), β_h is a k-dimensional vector of individual specific tastes for characteristics and marketing mix responsiveness, ξ_{jt} is unmeasured product characteristics that are unobserved by the researchers but considered by consumers in their purchase decisions and by marketers in their decision making, and ε_{hjt} is an i.i.d. random shock with a Type-I Extreme Value distribution. Consumer preferences are heterogeneous and to capture this, we model the taste vector β_h as a random draw from a multivariate normal distribution $N(\overline{\beta}, \Omega)$:

$$\beta_h = \overline{\beta} + \Omega^{1/2}\eta_h, \quad \eta_h \sim N(0, I_k), \tag{2}$$

where $\mathbf{\Omega}^{1/2}$ is the lower-triangular Cholesky factor of $\mathbf{\Omega}$, $\overline{\beta}$ is the mean parameter of the distribution of heterogeneity, and $\mathbf{\Omega}^{1/2}\eta_h$ is individual-specific deviation from the mean.

We can raise two issues related to ξ_{jt}. The first issue is the endogeneity problem. If marketers make their decisions based on the values of ξ_{jt}, marketing mix variables in x_{jt} would be correlated with ξ_{jt}. In particular, empirical research has repeatedly reported the correlation between price and ξ_{jt} (or price endogeneity) in disaggregate as well as in aggregate data. Due to this correlation, ξ_{jt} is not necessarily mean zero given x_{jt} and thus, we cannot treat it as another error component and integrate it out of the demand function. Second, regardless of the correlation with x_{jt}, ignoring ξ_{jt} would force the model to absorb these effects in the i.i.d. random shock and/or the explained part of the utility $x'_{jt}\beta_h$. As a result, one could get biased estimates of model parameters.

Following Heckman (1978) and Hausman (1978), we explicitly introduce the endogeneity issue into the random coefficient logit model by the following specification:

$$x_{jt} = (I_k \otimes z'_{jt})\gamma_j + \nu_{jt}, \quad \nu_{jt} \sim i.i.d.\, N(0, \Sigma_{\nu_j}), \tag{3}$$

$$\xi_{jt} \sim i.i.d.\, N(0, \sigma^2_{\xi_j}), \tag{4}$$

$$Cov(\nu_{jt}, \xi_{jt}) = \lambda_j, \tag{5}$$

$$Cov(z_{jt}, \xi_{jt}) = 0 \ \ \forall t, \tag{6}$$

where z_{jt} is an L-dimensional vector of instrumental variables uncorrelated with ξ_{jt} but correlated with x_{jt}. z_{jt} includes exogenous variables in x_{jt}. The distributional assumptions in (3) and (4) allow us to directly apply SML estimation, as described next. Without loss of generality, we apply the Cholesky decomposition of the covariance matrix of $[\nu'_{jt}\ \xi'_{jt}]'$ in order to rewrite it as a function of two independent shocks:

$$\begin{bmatrix} \nu_{jt} \\ \xi_{jt} \end{bmatrix} = \begin{bmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{bmatrix} \begin{bmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{bmatrix}, \quad \begin{bmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{bmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_k & 0 \\ 0 & 1 \end{pmatrix} \right), \tag{7}$$

$$Cov(\omega_{1,it}, \omega_{1,jt}) = 0 \quad for \quad i \neq j,$$

$$Cov(\omega_{2,it}, \omega_{2,jt}) = 0 \quad for \quad i \neq j,$$

where $b_{11,j} = \Sigma^{1/2}_{\nu_j}$. Using (7), we can rewrite (1) and (3) as follows:

$$u_{hjt} = x'_{jt}\beta_h + b_{21,j}\omega_{1,jt} + b_{22,j}\omega_{2,jt} + \varepsilon_{hjt}, \tag{8}$$

$$x_{jt} = (I_k \otimes z'_{jt})\gamma_j + b_{11,j}\omega_{1,jt}, \tag{9}$$

Solving (9) for $\omega_{1,jt}$ and substituting this in (8) results in the following transformation of (1):

$$u_{hjt} = x'_{jt}\beta_h + \theta'_j(x_{jt} - (I_k \otimes z'_{jt})\gamma_j) + b_{22,j}\omega_{2,jt} + \varepsilon_{hjt}\,, \tag{10}$$

where $\theta'_j = b_{21,j}b_{11,j}^{-1}$. Recall that ε_{hjt} is an i.i.d. random shock with a Type-I Extreme Value distribution and $\omega_{2,jt}$ is an i.i.d. random shock with a standard normal distribution. More importantly, $\omega_{2,jt}$ is uncorrelated with any other term in (10); we refer to this term as the exogenous unmeasured product characteristic or EUPC. In our approach we treat $\omega_{2,jt}$ as an additional error component and integrate it out of the demand function. Also note that $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$ is the residual from the regression in (3) and works as a bias correction term.

Our model is closely related to the "control function" method of Petrin and Train (2004). This method is based on a decomposition of the unmeasured product characteristic term as follows: $\xi_{jt} = f_j(\nu_t) + \varpi_{jt}$ where $\nu_t = [\nu'_{1t} \cdots \nu'_{Jt}]'$. The "control function" $f_j(\nu_t)$ is a function of the residuals ν_t obtained from the first stage regression in (3); this term controls for endogeneity. The new error components ϖ_{jt} are similar to our EUPC $\omega_{2,jt}$. While the proposed method assumes joint-normality of ν_{jt} and ξ_{jt}, the control function method requires specification of the functional form of $f_j(.)$ and the distribution of ϖ_{jt}. Note that we can get (10) from Petrin and Train's model by letting $f_j(\nu_t) = \theta'_j\nu_{jt}$ and $\varpi_{jt} = b_{22,j}\omega_{2,jt}$ along with the assumption of normality. In an empirical application of their model to the original automobile data of BLP (1995), Petrin and Train (2004) specified ϖ_{jt} to be Normal and hence this term was not separately identifiable from the normal random deviate in the constant term in utility. As a result the term did not need to be handled separately. However, in our model specification of brand choice which is popular in marketing, we need to separately integrate out the EUPC, as we explain in the next subsection.

Estimation procedures and endogeneity tests

From (2) and (10) we can derive the logit probability that consumer *h* chooses alternative *j*:

$$P_{hjt} = \frac{\exp(x'_{jt}\overline{\beta} + \theta'_j(x_{jt} - (I_k \otimes z'_{jt})\gamma_j) + b_{22,j}\omega_{2,jt} + x'_{jt}\Omega^{1/2}\eta_h)}{1 + \sum_{i=1}^{J} \exp(x'_{it}\overline{\beta} + \theta'_i(x_{it} - (I_k \otimes z'_{it})\gamma_i) + b_{22,i}\omega_{2,it} + x'_{it}\Omega^{1/2}\eta_h)}, \quad j=1,\ldots,J+1. \quad (11)$$

(11) has the usual random coefficient logit form except a bias correction term $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$ and time- and alternative-specific shocks $\omega_{2,t} = [\omega_{2,1t} \ldots \omega_{2,Jt}]$. Now we will describe a way to handle these shocks in the estimation. For expositional convenience, we first assume that the bias correction term, $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$, is given. Conditional on $\omega_{2,t}$, we can write the likelihood of the observed aggregate data in week t:[4] (Note that the assumption of a multinomial sampling process is made here, resulting in sampling error.)

$$L_{1,t}(\omega_{2,t}) = \left(\frac{H!}{n_{0t}!\cdots n_{J+1t}!}\right)\prod_{j=1}^{J+1}\left(\int P_{hjt}(\omega_{2,jt},\eta_h)\,\phi(\eta)\,d\eta\right)^{n_{jt}}, \quad (12)$$

where n_{jt} is the count of purchase trips for brand j in week t and $\phi(.)$ is the standard normal density function. Since $\omega_{2,t}$ are unknown, we again integrate them out:

$$L_{1,t} = \int L_{1,t}(\omega_{2,t})\phi(\omega_2)\,d\omega_2, \quad (13)$$

and the likelihood function for the sample of T weeks is

$$L_1 = \prod_{t=1}^{T} L_{1,t}. \quad (14)$$

For the computation of (12) and (13), we can use Monte Carlo simulation methods or SML (see Keane 1993).

In the implementation of SML, evaluation of the likelihood may encounter computational difficulties when n_{jt} is "large", because $\left(\int P_{hjt}(\omega_{2,jt},\eta_h)\,\phi(\eta)\,d\eta\right)^{n_{jt}}$ in (12) reaches machine zero fairly quickly. Although we were able to apply SML without this computational problem in our empirical application to aggregate data based on paper towel purchases of 880 households (discussed in a subsequent section), the problem is inescapable when H is large. The incidence of this computational problem depends on the size of H (or n_{jt}), distributions of choice probabilities, and the definition of machine zero on the particular computer and language used for estimation. Our approach to circumvent this problem, when it occurs, is to represent H consumers with a sample of tractable size, R. We use the observed sales in each time period t to compute shares of each of the $J+1$ products. We then draw a multinomial sample of size $R < H$ from these shares. This new, smaller sample is used to compute the likelihood function and obtain SML estimates[5].

A natural question arises regarding the new sample size: What is the optimal R? As R increases, we may expect efficiency gain. However, potential numerical inaccuracy also increases due to the increased exponent in (12). By trial and error with many different values of R ranging from 50 to 500 in simulation experiments, we determined that we get highly satisfactory results with R=100 but also note that the results do not change much with R. We use R=100 in all our simulation studies in the next section. We also applied the proposed method to many datasets generated from different values of H ranging from 1,000 to 100,000 and get satisfactory results in all cases. As H becomes larger, estimates are distributed closer to the true values but only marginally.

So far, we have assumed that the bias correction term, $(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)$, is given. However, γ_j needs to be estimated by maximizing the following likelihood function derived from (9):

$$L_{2,j} = (2\pi)^{-k/2} \left|\Sigma_{\nu_j}\right|^{-1/2} \exp\left(-0.5(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)'\Sigma_{\nu_j}^{-1}(x_{jt} - (I_k \otimes z'_{jt})\gamma_j)\right). \quad (15)$$

A joint estimation of the model can be performed by maximizing the following log likelihood function:

$$\ln L = \ln L_1 + \ln L_{2,1} + \cdots + \ln L_{2,J} . \quad (16)$$

In the above setting, a test for endogeneity is easy to perform. Note that $b_{21,j}$ captures the correlation between ξ_{jt} and x_{jt}. If there is no correlation between ξ_{jt} and x_{jt}, then $b_{21,j} = 0$ and $\theta'_j = b_{21,j} b_{11,j}^{-1} = 0$. Since our estimation procedure is based on SML, we can apply the standard hypothesis testing framework of MLE.[6] The null hypothesis (i.e. no endogeneity) is H_0 : $\theta = [\theta'_1 \;\; \cdots \;\; \theta'_J]' = 0$. The likelihood ratio (LR) test statistic and the Wald statistic can be derived as follows:

$$Wald = \hat{\theta}' Cov(\hat{\theta})^{-1} \hat{\theta} \sim \chi^2(Q),$$

$$LR = -2(\ln L_R - \ln L_{UR}) \sim \chi^2(Q),$$

where $\ln L_R$ and $\ln L_{UR}$ are the log likelihood value with- and without-restriction, respectively, and Q is the dimension of θ. More simply, we can obtain the significance of θ directly from the estimation result. We can regard this test as an extension of a regression-based Hausman test (Hausman 1978, 1983; also see Wooldridge 2001, p.118) or Wu test (Wu 1973) to a random coefficient logit model.

Implications of the Assumption of Joint Normality of ξ and ν[7]

The assumption of joint normality of unmeasured product characteristics ξ and price residuals ν in the proposed method (equations (3) and (4)), while standard from a statistical perspective, has strong economic implications. In particular, when the endogeneous explanatory variables are prices, this assumption is inconsistent with many forms of pricing behavior.

To explain this issue we begin with an example where the normality assumption *is consistent* with pricing. Rewriting the utility function in (1) by redefining x_{jt} to contain only non-price observed attributes, we have

$$u_{hjt} = \alpha_h p_{jt} + x'_{jt} \beta_h + \xi_{jt} + \varepsilon_{hjt}$$

where p_{jt} is the price of product j at t. Let the marginal cost of product j be linear in the observed and unobserved non-price attributes plus an error representing unobserved cost shocks:

$$MC_{jt} = x_{jt}^{'}\gamma + \lambda\xi_{jt} + \zeta_{jt}$$

Suppose that each product is priced at marginal cost, as in perfect competition. Then the price equation becomes

$$\begin{aligned} p_{jt} &= x_{jt}^{'}\gamma + \lambda\xi_{jt} + \zeta_{jt} \\ &= x_{jt}^{'}\gamma + \nu_{jt} \end{aligned}$$

In this situation the assumption of a normal distribution for unmeasured product characteristics ξ and a normal distribution for the marginal cost shocks ζ implies a joint normal for the error in the pricing equation ν and the unmeasured product characteristics ξ. This is also the case when prices are equal to marginal cost plus a fixed markup. Other forms of pricing do not yield this result. For example, under two prominent theories of pricing -- monopoly pricing and Nash pricing in a differentiated products oligopoly -- prices are some markup over marginal cost, where the markup depends on elasticities of demand at the prices. The pricing equation is

$$p_{jt} = x_{jt}^{'}\gamma + MK_{jt}(p, x, \xi) + \lambda\xi_{jt} + \zeta_{jt}$$

where *MK* denotes the profit maximizing markup. There is no way that this pricing equation can be neatly expressed in the form $p_{jt} = x_{jt}^{'}\tau + \nu_{jt}$ with a normal distribution for ν. The distribution of ν is defined implicitly by the solution to the pricing equation which has prices on both sides. It is not a simple task to derive a distribution for ν from assumed distributions for ξ and ζ. Even if we derive the distribution, it is not guaranteed to be normal. Furthermore, the distribution of ν will not be independent of *x*.

The foregoing discussion shows that the normality assumption is not inconsequential. Two factors mitigate the severity of the consequences in practice. First, the cost plus fixed markup model of pricing is widely practiced; Shim and Sudit (1995) report that it is used by over 80% of managers at manufacturing firms. Second, we find in our simulation studies that demand estimates from the proposed model are quite robust to violations of distributional assumptions.

Simulation study

We conduct simulation experiments with the following goals: a) to assess the performance of the proposed SML method; b) to assess robustness of the proposed method to key distributional assumptions about the unmeasured product characteristics, and c) to assess the implications of omitting the exogenous unmeasured product characteristics (EUPC) from the estimating model. To achieve these goals we consider five cases that are summarized in the table below.

Case	Data Generating Process	Goals
1.	Prices are endogenous, distribution of consumer heterogeneity is simple	Examine performance of proposed SML method
2.	Prices are endogenous, distribution of consumer heterogeneity is complex	Examine performance of proposed SML method
3.	Same as case 1, plus distribution of EUPC, ω_2, has autocorrelation.	Assess robustness of proposed SML method to misspecification of distribution of EUPC.
4.	Same as case 1, except that distributions of EUPC ω_2 and endogenous UPC ω_1 are uniform instead of normal.	Assess robustness of proposed SML method to violations of joint normality of ν and ξ.
5.	Same as case 1.	Assess implications of ignoring the EUPC in the estimation process.

In each of the five cases 100 datasets are generated as replicates, and models are estimated on each dataset to obtain the empirical sampling distribution of the parameter estimates. Data are generated from a sample of H=100,000 households.[8] We implement SML with R=100 as described in the previous section.

Case 1: Simple heterogeneity

We generate data by (1)-(6). The specific data generating process (DGP) and the parameter values we assign are summarized below:

$$u_{hjt} = x'_{jt}\beta_h + \xi_{jt} + \varepsilon_{hjt}, \quad u_{h(J+1)t} = \varepsilon_{h(J+1)t}, \text{ if no purchase}, \quad \varepsilon_{hjt} \sim i.i.d.\ EVI, \tag{17}$$

$$x_{jt} = [c_{1jt} \quad c_{2jt} \quad x_{1jt} \quad x_{2jt}]', \tag{18}$$

$$c_{1jt} = I(j=1), \quad c_{2jt} = I(j=2), \tag{19}$$

$$x_{1jt} = z_{1jt} + z_{2jt} + \omega_{1jt}, \quad x_{2jt} = I(\upsilon_{jt} > 0.7), \quad \upsilon_{jt} \sim Unif(0,1), \tag{20}$$

$$\xi_{jt} = \omega_{1jt} + \omega_{2jt}, \tag{21}$$

$$\omega_{1jt},\ \omega_{2jt},\ z_{1jt},\ z_{2jt} \sim i.i.d\ N(0, 0.5), \tag{22}$$

$$\beta_h = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \bar{\beta}_3 \\ \bar{\beta}_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h = \begin{bmatrix} 0.2 \\ 0.5 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h, \quad \eta_h \sim N(0, I_4), \tag{23}$$

where $I(.)$denotes an indicator function and I_k denotes a k-dimensional identity matrix. Note that $x_{1,jt}$ is correlated with ξ_{jt} through ω_{1jt} and this results in the endogeneity problem. We assume that individuals have heterogeneous tastes with respect to $x_{1,jt}$ only, J=2, T=50 or 100, and H=100,000. Even though utility is defined and generated at the disaggregate-level, we use only aggregate data for estimation. We have eight instrumental variables (c_{1jt}, $c_{1jt}z_{1jt}$, $c_{1jt}z_{2jt}$, $c_{1jt}x_{2jt}$, c_{2jt}, $c_{2jt}z_{1jt}$, $c_{2jt}z_{2jt}$, and $c_{2jt}x_{2jt}$) and five parameters to estimate ($\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\beta}_3$, $\bar{\beta}_4$, and σ_{33}). Hence the model is over-identified.

Insert Table 1 here

The first two moments of the empirical sampling distributions of the parameter estimates are summarized in Table 1. The proposed method works well in the recovery of mean and heterogeneity parameters. Even when the sample is as small as T=50, the estimates of mean and heterogeneity parameters are distributed close to the true values and we can conclude that the method provides unbiased estimates.

An additional feature of the proposed method is a simple test of endogeneity. By checking the significance of θ (= θ_1 = θ_2), we can formally test whether endogeneity is present. Note that in the present DGP, x_{1jt} is correlated with ξ_{jt} (i.e., endogeneity is present). In Table 1 we observe that the empirical distribution of θ is tightly distributed around the true value 1 for both T=50 and T=100 leading to the correct conclusion that θ is significantly different from zero and that endogeneity is present.

Case 2: Full heterogeneity

Here we generate data using a more complete heterogeneity distribution. The specific DGP and the parameter values we assign are the same as in Case 1 except for (23). The modified heterogeneity distributions are expressed as follows:

$$\beta_h = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \bar{\beta}_3 \\ \bar{\beta}_4 \end{bmatrix} + \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix} \eta_h = \begin{bmatrix} 0.2 \\ 0.5 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \eta_h, \quad \eta_h \sim N(0, I_4). \quad (23a)$$

Once again, J=2, T=50 and 100, and H=100,000. Note that consumers have heterogeneous tastes with respect to all x_{jt}. We have eight instruments (c_{1jt}, $c_{1jt}z_{1jt}$, $c_{1jt}z_{2jt}$, $c_{1jt}x_{2jt}$, c_{2jt}, $c_{2jt}z_{1jt}$, $c_{2jt}z_{2jt}$, and $c_{2jt}x_{2jt}$) and eight parameters to estimate ($\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\beta}_3$, $\bar{\beta}_4$, σ_{11}, σ_{22}, σ_{33}, and σ_{44}). Thus, the model is exactly identified.

The first two moments of the empirical distributions of the parameter estimates are summarized in Table 2. The estimates of mean parameters (i.e. $\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\beta}_3$, and $\bar{\beta}_4$) and heterogeneity parameters (i.e., σ_{11}, σ_{22}, σ_{33}, and σ_{44}) are close to their true values indicating that the method provides unbiased estimates.

Insert Table 2 here

However, dispersions of the distribution are larger than those in Case 1. This is due to the additional complexity in heterogeneity distributions of the current DGP. Note also that the empirical distribution of θ is tightly distributed around the true value 1 for both T=50 and T=100, supporting the presence of endogeneity in the data, which is the case in our DGP.

Case 3: Misspecification due to Autocorrelation in $\omega_{2,jt}$

In the DGP we assume autocorrelation in the distribution of the unmeasured product characteristic. All other details of the DGP remain the same as in Case 1 except (22) which is changed as follows:

$$\omega_{2jt} = \phi\,\omega_{2jt-1} + \varsigma_{jt} = 0.8\,\omega_{2jt-1} + \varsigma_{jt}\,, \tag{22a}$$

$$\varsigma_{jt} \sim i.i.d\ N(0,0.18)\,, \tag{22b}$$

In particular, we consider an AR(1) process with positive AR parameter, ϕ = 0.8. This implies that the unmeasured product characteristics are now positively correlated over time and the shock is rather persistent. From (22a), we can also see that $Var(\omega_{2jt}) = Var(\varsigma_{jt})/(1-\phi^2) = 0.5$. Time-series plots of $\omega_{2,jt}$'s (not shown for reasons of space) randomly generated from (22a)-(22b) confirm that $\omega_{2,jt}$ is highly autocorrelated. The estimation model for our proposed method remains the same as in Case 1, leading to misspecification.

Insert Table 3 here

Estimation results are shown in Table 3. We find that despite the misspecification estimates of all parameters are distributed around the true values, although dispersions of some parameters, particularly $\overline{\beta}_1$ and $\overline{\beta}_2$, are larger than in Case 1.

Case 4: Misspecification of Distribution of $\omega_{1,jt}$ and $\omega_{2,jt}$

As discussed, several pricing behaviors are inconsistent with the assumption of joint normality of the error in the pricing equation and the unmeasured product characteristics. In this study, we investigate the performance of the proposed method under non-Normal $\omega_{1,jt}$. In particular, we generate $\omega_{1,jt}$ and $\omega_{1,jt}$ from Uniform distributions. All other details of the DGP remain the same as in Case 1 except (22) which is changed as follows:

$$z_{1jt}, \quad z_{2jt} \sim i.i.d\ N(0, 0.5), \tag{22c}$$

$$\omega_{1jt}, \quad \omega_{2jt} \sim Unif(-1,1), \tag{22d}$$

The estimation model for our proposed method remains the same as in Case 1, leading to misspecification.

Insert Table 4 here

Estimation results are shown in Table 4. Essentially, all the favorable results for the proposed model that were obtained in Cases 1-3 are retained under this form of misspecification.

Case 5: Omitted Exogenous Unmeasured Product Characteristics

We examine the implications of ignoring unmeasured product characteristics even when they do not create an endogeneity issue. Chintagunta, Dubé, and Goh (2005) found that such omission led to higher estimated taste dispersion. However, we believe (and explain subsequently) that it is also possible that such misspecification would create biases in mean parameters. The specific data generating process (DGP) and the parameter values we assign are summarized below:

$$u_{hjt} = x'_{jt}\beta_h + \omega_{jt} + \varepsilon_{hjt}, \quad u_{h(J+1)t} = \varepsilon_{h(J+1)t}, \text{ if no purchase,} \quad \varepsilon_{hjt} \sim i.i.d.\ EVI, \tag{24}$$

$$x_{jt} = [c_{1jt} \quad c_{2jt} \quad x_{1jt} \quad x_{2jt}]',$$

$$c_{1jt} = I(j=1), \quad c_{2jt} = I(j=2),$$

$$x_{1jt}, \ \omega_{jt} \sim i.i.d \ N(0,1),$$

$$x_{2jt} = I(\upsilon_{jt} > 0.7), \quad \upsilon_{jt} \sim Unif(0,1),$$

$$\beta_h = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \bar{\beta}_3 \\ \bar{\beta}_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h = \begin{bmatrix} 0.2 \\ 0.5 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \eta_h, \ \ \eta_h \sim N(0, I_4),$$

where *I(.)* denotes an indicator function and I_k denotes a *k*-dimensional identity matrix. Note that $x'_{jt}\beta_h$, ω_{jt} and ε_{hjt} are mutually uncorrelated. We assume that consumers are only heterogeneous in their taste for x_{1jt}, J=2, T=100 and 200, and H=100,000. We estimate the proposed model and a benchmark model that omits the exogenous unmeasured product characteristics, ω_{jt}. The reason we consider larger values of *T* in this simulation as compared with Cases 1-4 is to illustrate more precisely the biases this omission causes.

Insert Table 5 here

The first two moments of the empirical distributions for the parameter estimates are summarized in Table 5. The performance of the proposed method is similar to that in Cases 1-4. Estimates are tightly distributed around the true value. Considering the estimates when ω_{jt} is omitted from the estimation (see the relevant columns in Table 5), we see that when T=100, estimates of $\bar{\beta}_3, \bar{\beta}_4$, and σ_{33} are biased toward zero. In particular, the true value of $\bar{\beta}_3$ is outside its 1.645 standard-error confidence band obtained from the empirical distribution of the parameter estimates (i.e. 90% confidence band of normal distribution). When *T* becomes 200, the biases become more apparent. Now the true values for $\bar{\beta}_3$ and $\bar{\beta}_4$ are out of their 1.645 standard-error confidence bands obtained from the empirical distributions of the parameter estimates.

To explain these biases, let us rewrite equation (24) as $u_{hjt} = \mu_{hjt} + \omega_{hjt} + \varepsilon_{hjt}$. If the exogenous unmeasured product characteristics ω_{jt} are ignored in the estimation, μ_{hjt} will absorb some part of the variation in ω_{jt} and ε_{hjt} will absorb the rest. Consequently, we get $u_{hjt} = \tilde{\mu}_{hjt} + \tilde{\varepsilon}_{hjt}$ where $\tilde{\mu}_{hjt}$ and $\tilde{\varepsilon}_{hjt}$ are the new explained utility and unexplained utility, respectively, both inflated by the ignored ω_{jt}. Since $Var(\tilde{\mu}_{hjt}) \geq Var(\mu_{hjt})$, we can expect over-estimated mean parameters and/or heterogeneity parameters. Similarly, we can expect $Var(\tilde{\varepsilon}_{hjt}) \geq Var(\varepsilon_{hjt})$ but this influences the model parameters in a way rather different from that of $\tilde{\mu}_{hjt}$. Due to the logit specification, the model will regard $\tilde{\varepsilon}_{hjt}$ as a random shock with a Type-I Extreme Value distribution and normalize the utility according to $Var(\tilde{\varepsilon}_{hjt})$. Since $Var(\tilde{\varepsilon}_{hjt}) \geq Var(\varepsilon_{hjt})$, $\tilde{\mu}_{hjt}$ and u_{hjt} will be scaled down by the factor $Var(\varepsilon_{hjt})/Var(\tilde{\varepsilon}_{hjt})$, which is less than 1[9]. Consequently, we expect under-estimated mean parameters and/or heterogeneity parameters. Note that ignoring the EUPC influences the estimates of heterogeneity and mean parameters in two conflicting ways simultaneously: upward bias via $\tilde{\mu}_{hjt}$ and downward bias via $\tilde{\varepsilon}_{hjt}$. Since it is hard to predict which effect is greater, we cannot draw a general conclusion on the direction and severity of the biases.

In our simulated data it appears that the majority of the variance related to the ignored EUPC ω_{jt} is absorbed into the unexplained part of utility and consequently, the downward bias via $\tilde{\varepsilon}_{hjt}$ overwhelms the upward bias via $\tilde{\mu}_{hjt}$. Although not significant, we observe biases in the same direction in heterogeneity parameters and these can be interpreted similarly.

Empirical Application

Data

The data used in the study are histories of paper towel purchases of 880 households at an independent supermarket in Pittsburgh, Pennsylvania over 103 weeks through 1998 and 1999. The data are collected using a frequent shopper card. We include the four largest brands in the analysis: Bounty, Brawny, Scott, and Sparkle. The sales from these four major brands accounted for 77% of total category sales in our sample. Additionally, we include a “No purchase” option defined as shopping visits when none of the four brands of paper towels is purchased. Within each brand, the purchase of any one of different package sizes (i.e. number of rolls) was counted as a purchase of the brand. Price is defined on a per roll basis in our analysis. Price and promotion variables at the brand level were computed as market share-weighted averages of brand-size level variables. Descriptive statistics of the purchases, marketing mix variables, and wholesale prices, are provided in Table 6. About seven percent of store visits result in purchases of paper towels. Bounty is the dominant brand in the market, with over two-thirds market share.

Insert Table 6 here

To control for the endogeneity of price, we use weekly wholesale prices as instruments. (Following Kuksov and Villas-Boas (2007) and Chintagunta, Dubé, and Goh (2005), we assume that price is the only endogenous variable.) As expected, wholesale price is highly correlated with shelf price, with correlation coefficients ranging from 0.54 to 0.92 across the four brands. However, we do not expect the unmeasured product characteristics, especially those determined at retail (e.g. shelf space allocation) to be systematically related with wholesale prices. To the extent that this expectation is true, our instrumental variable is valid for controlling for the endogeneity of price. Brawny has the highest shelf and wholesale prices, followed by the largest brand Bounty. Brawny and Scott show high variances in both shelf prices and wholesale prices.

Sparkle, the lowest-priced alternative, shows much more frequent promotion than other brands.

Estimation and results

Although our interest in this paper is in the aggregate model, we first estimate disaggregate models to obtain a benchmark (recall that the data are available at the household level). Disaggregate data contain complete information while aggregate data lose some information due to aggregation. Therefore, we expect that estimates from disaggregate data are more reliable than the ones from aggregate data for the same model specification. For the estimation of disaggregate models we use a method proposed by Chintagunta, Dubé, and Goh (2005) that we term ML/IV and describe subsequently.

Next, we aggregate our data to weekly brand sales data to estimate the proposed model as well as two other aggregate models –RCL and OEUPC. RCL is an aggregate version of the usual Random Coefficient Logit model. Therefore, it assumes that there is neither endogeneity of price nor unmeasured product characteristics. Model OEUPC omits the exogenous unmeasured product characteristics, $\omega_{2,jt}$, as in simulation Case 5.

Using disaggregate data, we estimate the ML/IV model, which is a generalized two-step estimator suggested by Chintagunta, Dubé, and Goh (2005): 1) first, estimate the usual random coefficient logit model treating $\delta_{jt} = x'_{jt}\overline{\beta} + \xi_{jt}$ as fixed-effects and get estimates of fixed effects $\hat{\delta}_{jt}$ and two heterogeneity parameters, σ_{price} and $\sigma_{promotion}$ [10]; 2) second, apply instrumental variable technique to $\hat{\delta}_{jt}$ considering the estimation error of the first step. Specifically, $\hat{\overline{\beta}}_{ML/IV} = (X'P_z\hat{\vartheta}^{-1}P_zX)^{-1}X'P_z\hat{\vartheta}^{-1}P_z\hat{\delta}$ where $P_z = Z(Z'Z)^{-1}Z'$ is a projection matrix of instrument variable matrix Z, and $\hat{\vartheta}$ is covariance matrix of $\hat{\delta}$ from the first-step estimation. Using the disaggregate data we also estimate a generalized least-square estimator ML/LS:

$\hat{\bar{\beta}}_{ML/LS} = (X'\hat{\vartheta}^{-1}X)^{-1}X'\hat{\vartheta}^{-1}\hat{\delta}$. ML/LS does not consider endogeneity of price and thus, we can investigate the bias due to endogeneity by comparing ML/IV and ML/LS.

Table 7 reports estimation results of three models with aggregate data (the proposed model, RCL, and OEUPC) and two models with disaggregate data (ML/IV and ML/LS).

Insert Table 7 here

The disaggregate data results show a significant decrease in the absolute value of the price coefficient after using an instrumental variable. A Hausman test for price endogeneity rejects the null hypothesis of no endogeneity (i.e., $\bar{\beta}_{ML/LS} = \bar{\beta}_{ML/IV}$). The upward bias due to ignoring the endogeneity of price is contrary to typical previous findings of a downward bias (e.g., Besanko, G upta, and Jain 1998; Chintagunta, Dubé, and Goh 2005, and others). In the literature, the unmeasured product characteristics are usually believed to be positively correlated with price, which is consistent with the commonly reported downward bias in the price coefficient when this correlation is ignored. However, our result implies a negative correlation between price and unmeasured product characteristics. One explanation for this negative correlation lies in expanded shelf space allocation or favorable shelf locations of price-promoted products, activities that are unobserved in our data. If shelf space changes are a dominant component of the unmeasured product characteristics, we can expect these characteristics to be negatively correlated with prices.

Turning next to results from aggregate models, all estimates from the RCL model are significantly different from those from ML/IV, which may be considered closest to the true parameter values. These "biases" are due to the omission of the unmeasured product characteristics as well as the ignored endogeneity. OEUPC estimates tell us what happens if we ignore the exogenous unmeasured product characteristics. Estimates of "Sparkle", "Promotion", and "$\sigma_{promotion}$" are smaller in absolute value than the estimates from ML/IV. This pattern is similar to what we observed in Case 5. We may conclude that the majority of the variance related to the ignored exogenous unmeasured product characteristics $\omega_{2,jt}$ is absorbed into the unexplained part of utility and consequently, the downward bias via $\tilde{\varepsilon}_{hjt}$ overwhelms the upward bias via $\tilde{\mu}_{hjt}$.

Estimates from the proposed method are all reasonably close to those of ML/IV and formally, they are not significantly different at a 5% significance level. We observe that standard errors of the proposed method are on average approximately 5 times those of ML/IV. This lower efficiency can be attributed to the information loss due to data aggregation. Estimate of "$\sigma_{promotion}$" is significantly different from that of OEUPC. Moreover, estimates of $b_{22,Bounty}$, $b_{22,Brawny}$, $b_{22,Sparkle}$, and $b_{22,Scott}$ are all significant. These results imply that EUPC needs to be accommodated appropriately in the estimation.

By examining estimates of θ's, we can easily perform a formal test of endogeneity. Furthermore, the estimate indicates the sign of correlation between the unmeasured product characteristics and price.[11] We obtained highly significant negative estimates of θ_{Bounty}, θ_{Brawny}, and $\theta_{Sparkle}$, indicating that the endogeneity problem does exist in Bounty, Brawny, and Sparkles, and that the unmeasured product characteristics are negatively correlated with price. This confirms findings from the disaggregate models. The estimate of θ_{Scott} is not significantly different from zero and thus, we cannot reject the null hypothesis of no endogeneity in Scott. Using the previous display and shelf space allocation argument, we may attribute this to a different shelf allocation practice of the retailer with respect to Scott.

Conclusion

In this paper, we propose a Simulated Maximum Likelihood estimation method for the random coefficient logit model using aggregate data, accounting for heterogeneity and endogeneity. Our approach is suitable when observed brand shares contain sampling error. We show in simulated data that the proposed method provides unbiased and efficient estimates of demand parameters. Further methodological advantages of the proposed method include: 1) the proposed method provides endogeneity test statistics as a by-product; 2) it directly provides the direction of endogeneity bias, or the sign of correlation between endogenous regressors and the unmeasured product characteristics; and 3) the proposed method can be extended to incorporate Markov regime-switching dynamics in parameters and is open to other extensions based on ML.

Substantively, we also provide a more complete picture of the problems related to the omission of unmeasured product characteristics. We have shown that, in addition to the endogeneity problem, the omission can cause downward or upward biases in the estimates of mean parameters and/or heterogeneity parameters in the random coefficient logit model. Previously, Chintagunta, Dubé, and Goh (2005) noted an upward bias in heterogeneity parameters due to this omission and confirmed this result using scanner panel data on margarine purchases. In this paper, we identify the possibility of other biases. As an example, in simulation Case 5 we found that downward biases in mean parameters can result from an omission of the unmeasured product characteristics. This result was also found in an empirical application to paper towels data.

In the paper towel data, we also found a negative correlation between the unmeasured product characteristics and prices, a result that was confirmed by the disaggregate data. This finding is new to the literature. Furthermore, so far the correlation between price and the unmeasured product characteristics has been indirectly inferred from the direction of the endogeneity bias rather than directly estimated as in our proposed method.

The most important limitation of our proposed method is that the assumption of joint normality of the unmeasured product characteristics and the error in the pricing equation is inconsistent with a number of pricing behaviors. However, simulation experiments showed that misspecification of the distribution does not hamper performance of the proposed method severely. We should also note that the proposed method is based on SML which is equivalent to ML when sufficient draws are used for the numerical integration. An attractive property of MLEs is that even under model misspecification, MLEs are strongly consistent in that they minimize the Kullback-Leibler Information Criterion. While this attractive property of MLEs alleviates our concerns somewhat, more rigorous study is needed with respect to misspecification of the unmeasured product characteristics.

The proposed method does not explicitly model serial correlation in the unmeasured product characteristics. However, simulation experiments showed that the autocorrelation in the unmeasured product characteristics does not hamper performance of the proposed method severely. We expect that a large part of this serial correlation can be captured by allowing Markov regime-switching dynamics in brand specific constants. Otherwise, we can extend the proposed method to explicitly model serial correlation by incorporating ARMA models. Extensions in this direction may also provide us with a deeper understanding of the unmeasured product characteristics.

References

Albuquerque, Paulo and Bart. J. Bronnenberg (2006), “Measuring Consumer Switching to a New Brand Across Local Markets,” working paper, University of Rochester, Rochester, NY.

Berry, Steven T. (1994), “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25 (2), 242–62.

Berry, Steven T., James Levinsohn, and Ariel Pakes (2004), “ Differentiated Products Demand System from a Combination of Micro and Macro Data: The New Car Market,” *Journal of Political Economy*, 112 (1), 68–105.

———, ———, and ——— (1995), “Automobile Prices in Market Equilibrium.,” *Econometrica,* 63 (4), 841–90.

———, O.B. Linton, and Ariel Pakes (2004), “Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems, “*Review of Economic Studies*,” 71 (248), 613–54.

Besanko, David, Sachin Gupta, and Dipak C. Jain (1998), “Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework,” *Management Scienc*e, 44 (November), 1533–47.

Bodapati, Anand and Sachin Gupta (2004), “The Recoverability of Segmentation Structure from Store-Level Aggregate Data,” *Journal of Marketing Research*, 41 (3), 351–64.

Bronnenberg, Bart J. and Vijay Mahajan (2001), “Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables,” *Marketing Science*, 20 (3), 284–99.

Chen, Yuxin and ShaYang (2007), “Estimating Disaggregate Models Using Aggregate Data Through Augmentation of Individual Choice,” *Journal of Marketing Research*, 44 (4), 613–21.

Chintagunta, Pradeep K. (2001), “Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data,” *Marketing Science*, 20 (4), 442–456.

———, Jean-Pierre Dubé and Khim Yong Goh (2005), “Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Model,” *Management Science*, 51 (5), 832–49.

Draganska, Mikhaila and Dipak C. Jain (2002), “A Likelihood Approach to Estimating Market Equilibrium Models,” *Management Science*, 50(5), 605–616.

Goolsbee, Austen and Amil Petrin (2004), “The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV,” *Econometrica*, 72 (2), 351–81.

Hausman, Jerry A. (1978), “Specification Tests in Econometrics,” *Econometrica*, 46 (6), 1251–72.

——— (1983), “Specification and Estimation of Simultaneous Equation Models,” in *Handbook of Econometrics*, Vol. 1, Z. Griliches and M. Intriligator, eds. Amsterdam, North Holland.

Heckman, James (1978), “Dummy Endogenous Variables in a Simultaneous Equation System,” *Econometrica*, 46 (4), 931–59.

Jiang, Renna, Puneet Manchanda, and Peter.E.Rossi (2007), “Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data,” working paper, University of Chicago, Graduate School of Business.

Keane, Michael P. (1993), “Simulation Estimation for Panel Data Models with Limited Dependent Variables,” in *The Handbook of Statistics & Econometrics*, C.R. Rao, G.S. Maddala, D. Vinod, eds. Elsevier Science Publishers, Amsterdam, Netherlands, 545–72.

——— (1997), “Modeling Heterogeneity and State Dependence in Consumer Choice Behavior,” *Journal of Business and Economic Statistics*, 15 (3), 310–27.

Kuksov, Dimitri and J. Miguel Villas-Boas (2007), “Endogeneity and Individual Consumer Choice,” *Journal of Marketing Research*, forthcoming.

Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta (2004), “Response Modeling with Nonrandom Marketing Mix Variables,” *Journal of Marketing Research*, 41 (2), 467–78.

Musalem, Andres, Eric. T. Bradlow, and Jagmohan S. Raju (2007), "Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data," *Journal of Applied Econometrics*, forthcoming.

Petrin, Amil (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Econom*y, 110 (4), 705–729.

——— and Kenneth Train (2004), "Omitted Product Attributes in Discrete Choice Models," working paper, University of Chicago, Chicago, IL.

Shim, Eunsup and Ephraim F. Sudit (1995), "How Manufacturers Price Products," *Management Accounting*, 76 (8), 37–39.

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

——— (2001 b), "Structural Analysis of Competitive Pricing in the Presence of a Strategic Retailer," *Marketing Science*, 20(3), 244-64.

Train, Kenneth (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK.

Villas-Boas, J. Miguel and Russell Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45 (10), 1324–38.

Villas-Boas, J. Miguel and Ying Zhao (2005), "Retailer, Manufacturers, and Individual Consumers: Modeling the Supply Side in the Ketchup Marketplace," *Journal of Marketing Research*, 42 (1), 83–95.

White, Halbert (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50 (1), 1–25.

Wooldridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Mass.: MIT Press.

Wu, De-Min (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41 (4), 733–50.

Yang, Sha, Yuxin Chen and Greg M. Allenby (2003), “Bayesian Analysis of Simultaneous Demand and Supply,” *Quantitative Marketing and Economics*, 1(3), 244–264.

Yatchew, Adonis and Zvi Griliches (1985), “Specification Error in Probit Models,” *The Review of Economics and Statistics*, 67, 134-139.

Footnotes

[1] In this paper, we refer to a model that uses aggregate- or market-level data as an aggregate model, and a model that uses household- or individual-level data as a disaggregate model. Within this definition, the aggregate model can specify utility at the household- or individual-level.

[2] We do not include here papers that tackle a related form of endogeneity in which marketing variables are set as a function of consumer responsiveness (e.g. Manchanda et al. 2004) or cross-sectional sales differences (e.g. Bronnenberg and Mahajan 2001).

[3] Say we consider J inside alternatives and T markets. Since the $J+1$th alternative (outside good) is normalized, degrees of freedom in aggregate data are $J \times T$, which is the number of fixed-effects to be estimated.

[4] For the rigorous derivation of this likelihood function, see Bodapati and Gupta (2004).

[5] Statistical properties of this estimator are provided in a Web-based appendix.

[6] If the number of draws in the simulation rises faster than the sample size, SML is consistent, asymptotically normal and efficient, and equivalent to ML (Train 2003 p.259), justifying our application of the standard hypothesis testing framework of ML.

[7] We are grateful to an anonymous reviewer for bringing this important issue to our attention. Our discussion in this sub-section borrows heavily from the reviewer's comments.

[8] All five simulations were also conducted with 1,000 households and the substantive findings were identical to those reported here for 100,000 households. Results are available from the authors on request.

[9] This argument is similar to *attenuation bias* mentioned by Yatchew and Griliches (1985) in a homogeneous probit model.

[10] We assume that individuals have heterogeneous tastes with respect to price and promotion only. Our data provide nine moment conditions at most and, given these, we cannot entertain more complicated specifications of the heterogeneity distribution.

[11] Recall that $\theta'_j = b_{21,j} b_{11,j}^{-1}$ and $Cov(\nu_{jt}, \xi_{jt}) = b_{21,j}$.

Table 1: Results of the Simulation Study Case 1 –
Simple Heterogeneity
(based on 100 replications)

Parameters	True Values	T=50		T=100	
		Mean	SE*	Mean	SE
$\overline{\beta}_1$	0.200	0.212	0.136	0.206	0.100
$\overline{\beta}_2$	0.500	0.496	0.157	0.489	0.105
$\overline{\beta}_3$	-1.000	-1.007	0.163	-0.977	0.147
$\overline{\beta}_4$	1.000	0.990	0.205	0.980	0.158
$\theta_1 = \theta_2$	1.000	1.060	0.185	0.980	0.117
σ_{33}	1.000	1.005	0.194	1.009	0.152
$SD(\omega_{2jt})$	0.707	0.718	0.087	0.725	0.065

*SE is the standard deviation of the empirical sampling distribution of the estimate (based on 100 replications).

Table 2: Results of the Simulation Study Case 2 – Full Heterogeneity (based on 100 replications)

Parameters	True Values	T=50		T=100	
		Mean	SE	Mean	SE
$\bar{\beta}_1$	0.200	0.187	0.219	0.227	0.137
$\bar{\beta}_2$	0.500	0.518	0.174	0.531	0.122
$\bar{\beta}_3$	-1.000	-1.010	0.231	-0.983	0.171
$\bar{\beta}_4$	1.000	1.019	0.336	0.966	0.215
$\theta_1 = \theta_2$	1.000	1.026	0.252	0.991	0.175
σ_{11}	1.000	1.005	0.637	0.903	0.481
σ_{22}	1.000	0.916	0.628	0.875	0.495
σ_{33}	1.000	1.033	0.266	0.993	0.190
σ_{44}	1.000	1.050	0.654	0.970	0.471
$SD(\omega_{2jt})$	0.707	0.700	0.152	0.689	0.102

Table 3: Results of the Simulation Study Case 3 –

Misspecification due to Autocorrelation in ω_{2jt}

(based on 100 replications)

Parameters	True Values	T=50		T=100	
		Mean	SE	Mean	SE
$\bar{\beta}_1$	0.200	0.191	0.315	0.227	0.195
$\bar{\beta}_2$	0.500	0.445	0.261	0.505	0.247
$\bar{\beta}_3$	-1.000	-1.019	0.160	-0.973	0.153
$\bar{\beta}_4$	1.000	1.044	0.188	0.969	0.147
$\theta_1 = \theta_2$	1.000	1.043	0.173	0.985	0.143
σ_{33}	1.000	1.045	0.224	0.994	0.136
$SD(\omega_{2jt})$	0.707	0.676	0.138	0.675	0.086

Table 4: Results of the Simulation Study Case 4 –

Misspecification of the Distribution of ω_{1jt} and ω_{2jt}

(based on 100 replications)

Parameters	True Values	T=50		T=100	
		Mean	SE	Mean	SE
$\bar{\beta}_1$	0.200	0.199	0.125	0.196	0.090
$\bar{\beta}_2$	0.500	0.505	0.128	0.494	0.089
$\bar{\beta}_3$	-1.000	-0.988	0.147	-1.009	0.127
$\bar{\beta}_4$	1.000	0.986	0.189	1.005	0.131
$\theta_1 = \theta_2$	1.000	1.022	0.155	1.007	0.117
σ_{33}	1.000	1.017	0.193	1.006	0.137
$SD(\omega_{2jt})$	-	0.568	0.061	0.582	0.047

Table 5: Result of the Simulation Study Case 5 –
Exogenous Unmeasured Product Characteristics
(based on 100 replications)

Parameters	True values	T=100				T=200			
		Proposed method		Omitted ω_{jt}		Proposed method		Omitted ω_{jt}	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE
$\bar{\beta}_1$	0.200	0.212	0.192	0.310	0.115	0.230	0.129	0.301	0.086
$\bar{\beta}_2$	0.500	0.495	0.170	0.551	0.099	0.471	0.128	0.540	0.073
$\bar{\beta}_3$	-1.000	-0.951	0.144	-0.790	0.113	-0.946	0.119	-0.817	0.100
$\bar{\beta}_4$	1.000	0.961	0.221	0.807	0.148	0.949	0.195	0.803	0.119
σ_{33}	1.000	0.984	0.190	0.808	0.213	1.004	0.169	0.839	0.167
$SD(\omega_{jt})$	1.000	1.056	0.099	-	-	1.059	0.068	-	-

Table 6: Descriptive statistics of the Paper Towel Data
(Number of households = 880, Number of weeks = 103, Number of trips = 60,393)

Brands	Number of Purchases	Shelf Price ($ per roll)		Wholesale Price ($ per roll)		Correlation between shelf price and wholesale price	Promotion (% of store weeks)
		Mean	SD	Mean	SD		
Bounty	4,348	1.47	0.17	1.27	0.17	0.90	3%
Brawny	781	1.56	0.34	1.37	0.31	0.92	11%
Scott	558	1.39	0.39	1.13	0.28	0.94	6%
Sparkle	778	1.00	0.13	0.79	0.10	0.54	29%

Table 7: Estimation Results of Paper Towel Data

	Disaggregate Data				Aggregate Data					
Model→	ML/LS		ML/IV		RCL		OEUPC		Proposed method	
Parameters	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Bounty	0.76	0.12	-0.25	0.10	2.70	0.37	0.03	0.39	-0.18	0.43
Brawny	-0.65	0.13	-2.25	0.11	0.98	0.36	-1.88	0.39	-2.16	0.44
Scott	-1.35	0.12	-2.62	0.10	0.20	0.34	-2.19	0.37	-2.47	0.41
Sparkle	-1.70	0.10	-3.85	0.08	-0.53	0.31	-2.97	0.34	-3.28	0.39
Price	-2.57	0.10	-1.98	0.08	-5.73	0.58	-2.66	0.50	-2.53	0.53
Promotion	1.22	0.19	1.55	0.14	1.12	0.16	1.25	0.10	1.00	0.52
σ_{Price}	1.26	0.03	1.26	0.03	2.51	0.29	1.38	0.24	1.44	0.25
$\sigma_{Promotion}$	2.11	0.14	2.11	0.14	1.49	0.22	0.31	0.32	1.50	0.64
$b_{22,Bounty}$									0.27	0.03
$b_{22,Brawny}$									0.74	0.07
$b_{22,Scott}$									0.42	0.07
$b_{22,Sparkle}$									0.72	0.09
θ_{Bounty}							-1.87	0.27	-1.64	0.41
θ_{Brawny}							-4.66	0.39	-4.45	0.49
θ_{Scott}							0.25	0.36	0.30	0.53
$\theta_{Sparkle}$							-6.39	0.34	-6.57	1.12
AIC					52,353		52,241		51,474	
BIC					52,386		52,291		51,542	
log-like.					-26,168		-26,108		-25,721	

We perform joint estimation for the proposed method as described in the estimation section. However, in the table, AIC, BIC, and log-likelihood of the proposed method are calculated from the log-likelihood of the logit probabilities in Eq (14) or ln L_1 for direct comparison with RCL. Similarly, we perform joint estimation for OEUPC but calculate AIC, BIC, and log-likelihood from the logit probabilities in ln L_1.

WEB APPENDIX

to

A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data

Sungho Park
sp393@cornell.edu

Sachin Gupta
sg248@cornell.edu

Johnson Graduate School of Management
Cornell University
Ithaca NY 14853

May 2008

Sungho Park is a doctoral student and Sachin Gupta is Henrietta Johnson Louis Professor of Management and Professor of Marketing, both at the Johnson Graduate School of Management, Cornell University, Sage Hall, Ithaca NY 14853.

In this Appendix we demonstrate some statistical properties of the simulated maximum likelihood estimator proposed in Park and Gupta (2008). As described in Park and Gupta (2008), our estimation approach involves drawing a new sample of size *R* (smaller than *H*) based on shares in the observed sample of size *H*, in order to overcome numerical difficulties associated with using the observed sample, if *H* is large. Thus, our sample of size *R* could be considered a "two-stage sample". The alternative is a sample of size *R* taken directly from the population (call this a "one-stage sample"). We know that standard results, such as consistency, apply to SML estimates obtained from a one-stage sample. We determine some of the properties of the two-stage sample by comparing a one-stage sample with a two-stage sample, both of size *R*. If the two samples are very similar, we expect standard results to apply to the two-stage sample as well.

In this Appendix we proceed as follows. First we define a two-stage sample. Next, we use simulations to examine the similarity between one-stage and two-stage samples. Finally, we directly compare SML estimates obtained from one-stage and two-stage samples for different values of *R* in simulated data.

<u>Definition of two-stage sample</u>

Define

$$P_{jt} = \int \frac{\exp(x_{jt}\beta_h + \xi_{jt})}{1 + \sum \exp(x_{it}\beta_h + \xi_{it})} \phi(\beta)\, d\beta \,.$$

Let us assume that we observe aggregate shares $\{S_{jt}\}(= N_{jt}/H)$ based on a sample of size *H* which is finite, and thus observed shares contain sampling error. $\{N_{jt}\}$ $(\Sigma_j N_{jt} = H)$ are outcomes of H multinomial draws from probabilities $\{P_{jt}\}$. When $\{N_{jt}\}$ are large, numerical problems prevent direct application of SML. In this case, we perform *R* multinomial draws (*R*<*H*) from shares $\{S_{jt}\}$ and get outcomes $\{M_{jt}\}$ ($\Sigma_j M_{jt} = R$, define $Q_{jt} = M_{jt}/R$). This is our two-stage sample. By using $\{M_{jt}\}$ instead of $\{N_{jt}\}$ to obtain SML estimates, we can circumvent the numerical difficulty of handling large exponents.

<u>Comparison of two-stage sample with one-stage sample</u>

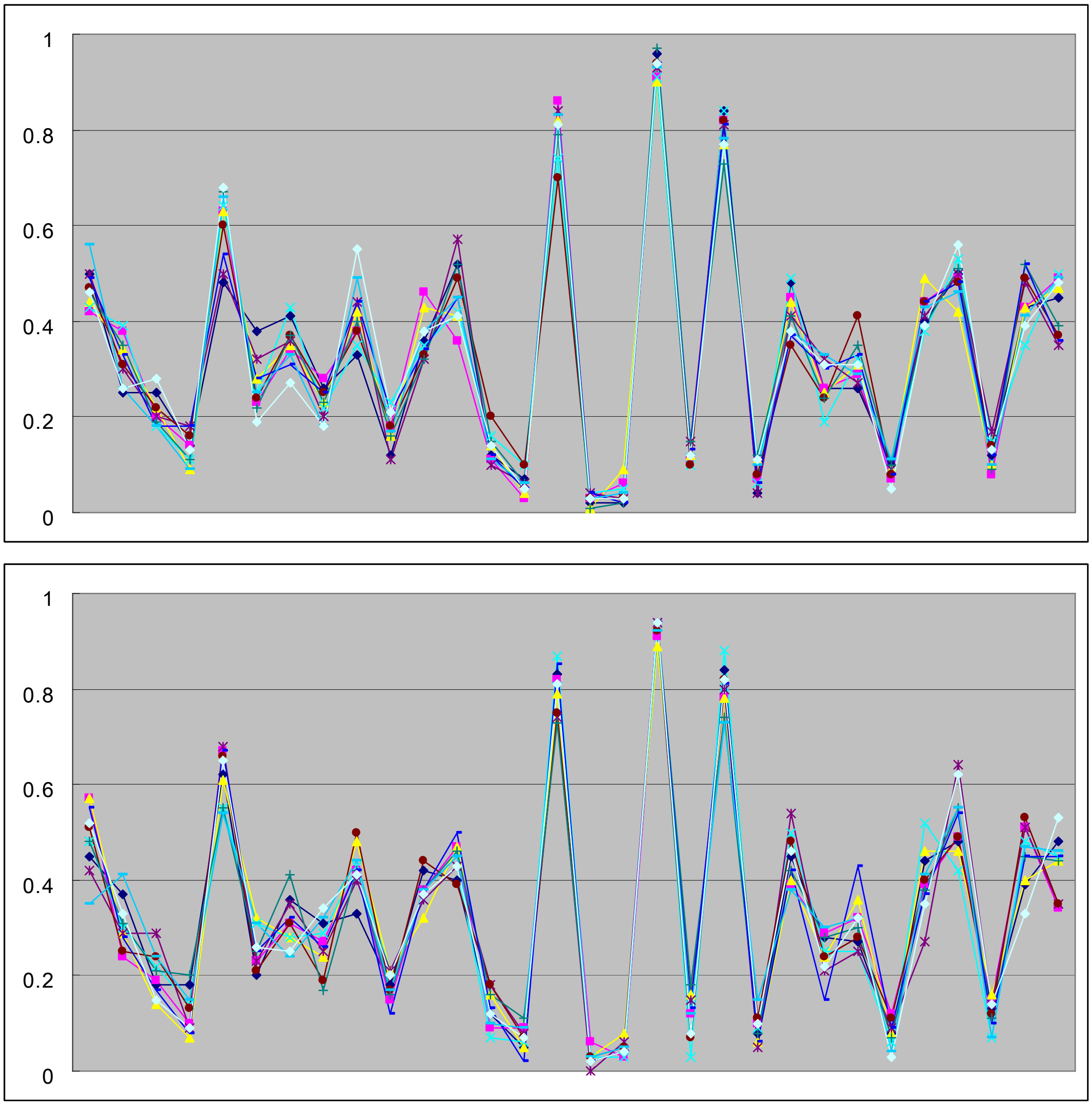
We use the DGP of case 1 of the simulation study reported in Park and Gupta (2008). We calculate $\{P_{jt}\}$ using 100,000 random draws and set *H*=1,000 and *R*=100. To generate a one-stage sample we perform *R*(=100) multinomial draws from $\{P_{jt}\}$. To generate a two-stage sample we perform *H*(=1,000) multinomial draws from $\{P_{jt}\}$, calculate $\{S_{jt}\}(=N_{jt}/H)$, and then perform *R*(=100) multinomial draws from $\{S_{jt}\}$ to obtain $\{M_{jt}\}$ or $\{Q_{jt}=M_{jt}/R\}$. For each set of $\{P_{jt}\}$, we generated 100 sets of one-stage samples and 100 sets of two-stage samples. So, for P_{jt} with each *j* and *t*, we have $Q_{jt,d}^{1-stage}$ (*d*=1,…,100) from one-stage sample and $Q_{jt,d}^{2-stage}$ (*d*=1,…,100) from two-stage sample.

In Figure W1 we show $Q_{jt,d}^{1-stage}$ and $Q_{jt,d}^{2-stage}$ for the first ten samples, (*d*=1,…,10). (We do not show all 100 samples to allow easier reading of the figure.) We find that both $Q_{jt,d}^{1-stage}$ and $Q_{jt,d}^{2-stage}$ are centered around $\{P_{jt}\}$ and their distributions look quite similar.

Figure W1. One-stage samples (top panel) vs. Two-stage samples (bottom panel)

x-axis: (t=1; j=1), (t=1; j=2), (t=1; j=3), (t=2; j=1), (t=2; j=2), (t=2; j=3),…

y-axis: $Q_{jt,d}^{1-stage}$ (top panel) and $Q_{jt,d}^{2-stage}$ (bottom panel) for d=1,…,10.



Next we calculate the average absolute distances of $Q_{jt,d}^{1-stage}$ and $Q_{jt,d}^{2-stage}$ respectively from P_{jt}.

We define

$$dist\_1-stage = \frac{\Sigma_{t=1\sim T}\Sigma_{j=1\sim J}(\Sigma_{d=1\sim 100}|P_{jt}-Q_{jt,d}^{1-stage}|/100)}{J*T}$$

$$dist\_2-stage = \frac{\Sigma_{t=1\sim T}\Sigma_{j=1\sim J}(\Sigma_{d=1\sim 100}|P_{jt}-Q_{jt,d}^{2-stage}|/100)}{J*T}$$

In our data $dist\_1-stage$ =0.033 and $dist\_2-stage$ =0.034. Table W1 summarizes the results from simulations with different values of *H* and *R*. In general, $dist\_1-stage$ is very close to $dist\_2-stage$. As *R* increases, both distances decrease. Comparing results for *H*=10,000 and *H*=1,000, we observe that $dist\_2-stage$ decreases only marginally as *H* increases. From all these results, we conclude that a one-stage sample is empirically equivalent to a two-stage sample and thus we expect that the standard results of SML apply to a two-stage sample.

Table W1. Results of Simulation Study

H	*R*	*dist_2-stage*	*dist_1-stage*
10,000	150	0.027	0.027
10,000	100	0.033	0.033
10,000	50	0.046	0.046
1,000	150	0.028	0.027
1,000	100	0.034	0.033
1,000	50	0.047	0.046

Statistical properties of the proposed SML estimator

We consider how SML estimates behave as *R* changes in a simulation study. We use the same DGP as in case 1 of the paper. We generate $\{N_{jt}\}$ with *H*=1,000 and then draw three sets of two-stage samples with *R*=(50, 100, 150) as well as one-stage samples of the same three sizes. We generated 200 data sets for this simulation study.

We expect that standard errors decrease as *R* increases. The rate of decrease is expected to be $\sqrt{N}$ if there is no numerical integration for heterogeneity and unmeasured product characteristics. In the presence of numerical integration, the rate is expected to be slower than

$\sqrt{N}$.

Table W2. Results of Simulation Study to Compare SML estimates from One-stage versus Two-stage Samples

Two-stage Sample							
Parameters	True values	H=1000, R=50		H=1000, R=100		H=1000, R=150	
		Mean	SE	Mean	SE	Mean	SE
$\overline{\beta}_1$	0.2	0.211	0.100	0.190	0.100	0.185	0.088
$\overline{\beta}_2$	0.5	0.515	0.102	0.512	0.095	0.492	0.090
$\overline{\beta}_3$	-1	-1.013	0.114	-0.976	0.105	-0.985	0.090
$\overline{\beta}_4$	1	0.988	0.144	0.980	0.122	0.991	0.093
$\theta_1 = \theta_2$	1	1.006	0.101	0.996	0.111	1.028	0.111
σ_{33}	1	1.011	0.133	1.000	0.122	1.005	0.109
$SD(\omega_{2jt})$	0.707	0.710	0.069	0.736	0.063	0.755	0.058

One-stage Sample							
Parameters	True values	R=50		R=100		R=150	
		Mean	SE	Mean	SE	Mean	SE
$\overline{\beta}_1$	0.2	0.206	0.095	0.191	0.092	0.196	0.089
$\overline{\beta}_2$	0.5	0.509	0.100	0.507	0.090	0.503	0.088
$\overline{\beta}_3$	-1	-0.984	0.110	-0.976	0.103	-0.980	0.101
$\overline{\beta}_4$	1	0.979	0.138	0.980	0.138	0.984	0.132
$\theta_1 = \theta_2$	1	0.994	0.106	0.990	0.104	0.994	0.102
σ_{33}	1	0.984	0.145	0.986	0.131	0.991	0.126
$SD(\omega_{2jt})$	0.707	0.687	0.063	0.699	0.059	0.714	0.058

The top panel in Table W2 summarizes the first two moments of estimates from two-stage samples. All estimates are tightly distributed around the true values. When we compare (*H*=1000, R=50) to (*H*=1000, *R*=100) and (*H*=1000, *R*=150), we observe that for each parameter (with the exception of □) SE's decrease as *R* increases. The rate of decrease is slower than $\sqrt{N}$.

The lower panel reports the first two moments of estimates from one-stage samples. Here we confirm all the same results as two-stage samples. By comparing upper and lower panels, we find that in all cases (*R*=50, *R*=100, and *R*=150), we obtain comparable results from one-stage and two-stage samples.

References

Park, Sungho, and Sachin Gupta (2008), "A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data," working paper, Johnson Graduate School of Management, Cornell University.