ENDOGENEITY AND INDIVIDUAL CONSUMER CHOICE*

DMITRI KUKSOV
(Washington University, St. Louis)

J. MIGUEL VILLAS-BOAS
(University of California, Berkeley)

Forthcoming, Journal of Marketing Research

October, 2007

* Comments by the Editor, the Associate Editor, and two anonymous reviewers are greatly appreciated. Address for correspondence: Haas School of Business, University of California at Berkeley, Berkeley, CA 94720-1900. E-mail: kuksov@olin.wustl.edu and villas@haas.berkeley.edu.

Endogeneity and Individual Consumer Choice

Abstract

In many markets one is likely to have both correlation across periods of the choices of one consumer and correlation across consumers of their choices in each period. The former is caused by consumer heterogeneity, while the latter may be the result of demand common shocks across consumers. Furthermore, if these common shocks are partially observed by the firms, then the market decisions of the firms end up being endogenous and correlated with the common shocks. Because the researcher cannot typically fully observe consumer heterogeneity and the common shocks, the estimation method has to account for the endogeneity of the decisions made by the firms. In this paper, we present a test for endogeneity under unobserved consumer heterogeneity and common shocks, which is based on a quasi-likelihood estimation method to consistently estimate the model parameters when endogenous firm behavior, unobserved heterogeneity, and common shocks are present. The test is a test of the differences in GMM coefficient estimates of a model with and without instrumenting for the explanatory variables. We show theoretically that in the estimation method we use, unobserved heterogeneity does not affect the consistency of the parameter estimates, but endogeneity, if not accounted for, may bias the results. We present estimation results from simulated and scanner panel data.

1. Introduction

Most markets seem to potentially have the following three features: (1) Consumers receive common demand shocks. (2) Firms observe, at least partially, these shocks and use these observations when deciding on prices and other variables that may affect demand. And (3) individual consumer preferences are stable, so that the choices of each individual consumer are correlated through time.

This paper presents a robust test of endogeneity when consumers receive common demand shocks, the choices of each individual consumer are correlated through time, and the researcher has panel data on consumer choices. The test is based on an estimation method of the consumer preference parameters when all the above three market features are present in a market.

Common demand shocks can cause problems in the estimation of the preference parameters when these shocks are unobservable to the researcher. For example, taste shocks, some advertising or coupon availability, changes of prices of competing goods, changes in the economic outlook or weather, can affect all consumers and may be unobservable to the researcher. In fact, if there were no common demand shocks (i.e., all shocks to demand were independent across consumers), then the aggregate demand from a large number of consumers standardized by the market potential could be obtained without error, which is empirically never true.

If these common demand shocks exist then it is quite possible that firms are able to observe these shocks, at least partially. In such a case, firms would use these observations in their decisions on price and other variables affecting demand, and therefore, these firm-controlled explanatory variables of the consumer choices would be endogenous. For example, if a firm finds that its demand received a positive shock, it might choose a higher price, as well as feature and display decisions, and the estimation method has to account for this possibility. This rules out, for example, standard discrete choice logit and probit estimation methods.[1]

At the same time, each consumer's decisions are correlated through time because the consumer preferences may be, at least to some extent, stable through the period under analysis. Some of this preference heterogeneity is unobserved to the researcher and has to be included in the econometric error term. This issue of unobserved consumer heterogeneity by itself has been the target of a substantial literature for more than a decade (e.g., Kamakura and Russell 1989, Chintagunta et al. 1991).[2]

[1] See, for example, Bass (1969) for an early analysis of endogeneity problems with aggregate data in the context of advertising and sales.

[2] See also, for example, Grover and Srinivasan (1992), Gonul and Srinivasan (1993), Rossi and Allenby (1993), Bucklin et al. (1995), Kim et al. (1999).

The test of endogeneity that we propose while accounting for both correlation across consumers in each period and correlation across periods from each consumer is based on an estimation method of simulated moments that accounts for these three features, where moments are derived from the quasi-likelihood function while instrumenting for the potentially endogenous explanatory variables of the consumer utility function.[3] Endogeneity can be tested by testing the statistical significance of the difference (up to a multiplicative constant) between the estimates of coefficients estimated with and without instrumenting. The method considers each consumer choice as one observation and includes the correlation across consumers in each period and the correlation across periods from each consumer in the estimates of the variance of the parameter estimates.

The need to instrument for price has been well established when estimating demand, where the prices and quantities result from a market equilibrium of supply and demand (see, e.g., Morgan 1990, Part 2, Ch. 6). The endogeneity problem in that setting has been also referred to as a simultaneity problem as it comes from a simultaneous setting of prices and quantities in the market. In that situation, the price correlates with the error terms because it is determined on the basis of quantities involving the realized error terms. Note, that for the endogeneity problem to occur, one does not need any strategic (i.e., not price taking) behavior by market participants, but only that the market price be correlated with the demand error term. Such a case occurs if firms know, at least partially, the demand error term when they are setting prices.

For a simple example illustrating the problem, consider a monopolist setting price p on a non-durable product, that is valued by consumer i as $v_i + \varepsilon_t$, where v_i, specific to each consumer i, is constant through time, and the random component ε_t is common across consumers, but varies with time. Suppose that v_i is distributed in the population with cumulative distribution function $F(v)$. Assume consumers have interest in a single unit of the product. Then the demand in period t is proportional to $1 - F(p_t - \varepsilon_t)$, which is a function of the difference between the price being set, p_t, and the variable disturbance ε_t. Suppose that the monopolist knows the state of demand, ε_t, before setting the price. Under general conditions we have that increases in ε_t lead to increases in price, and that these price increases are smaller than the increases in ε_t.[4] Thus, price and realized demand both move in the direction of ε_t. Hence, a researcher observing a series of realized prices

[3]The method provides consistent estimates of the consumer purchase behavior in the presence of heterogeneity, but does not estimate consumer heterogeneity per se. Explicit estimation of consumer heterogeneity may still be important, for example, for targeting purposes.

[4]Assuming that the marginal cost c is constant, consider the monopolist's problem: $\max_p (p-c)[1-F(p-\varepsilon)]$. For simplicity assume that $F(\cdot)$ is twice continuously differentiable, and the second order condition is satisfied with a strict inequality. Assume further that $F' + (p-c)F'' > 0$. The first order conditions imply that $1 - F - (p-c)F' = 0$. Totally differentiating with respect to ε and p, we obtain $\frac{dp}{d\varepsilon} = 1 - \frac{F'}{2F' + (p-c)F''}$, which means that $0 < \frac{dp}{d\varepsilon} < 1$.

and demands, and not correcting for the endogeneity, would incorrectly assume that the price has a positive value for the consumer.[5]

In the application below to scanner panel data on a ketchup market, we test for potential endogeneity in price, display, and feature. We find endogeneity problems in price, but not necessarily in display or feature.

In relation to the unobserved heterogeneity literature discussed above (e.g., Kamakura and Russell 1989, Chintagunta et al. 1991) this paper discusses how one can consider common shocks and endogeneity in addition to unobserved consumer heterogeneity.

Villas-Boas and Winer (1999) also consider the existence of endogeneity in prices in a model of consumer choice with panel data.[6] In relation to that paper, this paper has the following contributions. First, this paper models the possibility of unobserved consumer heterogeneity. Second, this paper presents an estimation method that is easier to implement, presents consistent estimates,[7] and can be realistically applied to choice sets with any number of alternatives. In particular, while the estimation methods presented in Villas-Boas and Winer required the computation of the multiple integral over the common shocks, this computation is no longer necessary here. This feature is important because it allows for a much easier applicability of endogeneity tests in choice model settings. Third, this paper considers endogeneity not only in price but in all the explanatory variables controlled by the firms. In the particular application below, in addition to price, we also consider the potential endogeneity of the feature and display variables. Finally, this paper considers the no purchase option and throughout the analysis, uses cost instruments, which may be seen as quite independent of the latent utility common shocks. Yang et al. (2003) present a Bayesian approach, suited for small samples, to accounting for price endogeneity, which includes the possibility of consumer unobserved heterogeneity.[8] In relation to that paper, this paper presents an estimation method that relies on fewer functional form assumptions, but requires a large sample to obtain

[5]Similar issues can be obtained under competition. For these issues see, for example, Moorthy (1985), Lal and Matutes (1994), Rao and Syam (2001).

[6]Villas-Boas and Zhao (2001) also consider price endogeneity while fully modeling the supply side of the market. For work on endogeneity issues with aggregate data (and not panel data) in similar models see also, for example, Berry et al. (1995), Besanko et al. (1998), Nevo (2001), Sudhir (2001). See also Roy et al. (1994) for the same type of analysis. For a general approach to estimation with panel data see Chamberlain (1984). See also Rivers and Vuong (1988), and Blundell and Powell (2004) for estimation methods related to the one presented here, and a previous version of this paper, Author(s)(2001).

[7]In Villas-Boas and Winer (1999), the results of simulated maximum likelihood estimation were only consistent if the number of draws in the simulation went to infinity at a faster rate than the sample size.

[8]The particular application does not consider latent utility re-scaling issues, uses lagged prices as "instruments" that, as noted in Villas-Boas and Winer (1999) (which also use those instruments in a part of the analysis), may lead to an under-estimation of the endogeneity bias, and does not acknowledge that the limited and full information specifications represent models with different functional forms.

statistical (asymptotic) results. Furthermore, this paper argues that, asymptotically, accounting for endogeneity is important in order to obtain consistent parameter estimates, while accounting for unobserved consumer heterogeneity is not (however, one still has to allow the random coefficients that result in the same variance between observations, i.e., it may be important to include the time-varying random coefficient terms). Note that this result is consistent with the meta-analytical results in Bijmolt et al. (2005) who show that while endogeneity does affect price elasticities, heterogeneity, on average, does not. Note also that this result is consistent with the results on aggregate data in Chintagunta (2001), that shows that both endogeneity and heterogeneity can create bias, since with aggregate data, modeling consumer heterogeneity (time-invariant random terms) is equivalent to having time varying random coefficients, and as noted above these are important to have if the true model has consumer heterogeneity. In independent work, Petrin and Train (2002), following up on Villas-Boas and Winer (1999), present results on a two-step "control function" estimation method (see also Blundell and Powell, 2004) similar to the one presented in this paper. Petrin and Train focus on the comparison of this approach with a fixed effects approach, while the point of this paper is, in addition of presenting a simpler estimation method than in Villas-Boas and Winer, to investigate the effects of common demand shocks and unobserved consumer heterogeneity on the estimation in a general framework. The relation with Petrin and Train is further noted in Section 3.2. Goolsbee and Petrin (2004) and Chintagunta et al. (2005) present the use of the fixed effects approach of aggregate models (e.g., Berry et al. 1995) to individual level data. Because the analysis in Chintagunta et al. (2005) uses panel data (in contrast to Goolsbee and Petrin) it can consider unobserved consumer heterogeneity, and obtain the maximum likelihood estimates of the heterogeneity parameters given the fixed effects. Although the estimation method considered here accounts for both unobserved consumer heterogeneity and common demand shocks and is simple to implement, there may be loss of efficiency in the estimation procedure as it is based on a quasi-likelihood function instead of the true likelihood, and on the method of simulated moments, which may bring additional error into the estimation of the parameters.

The remainder of the paper is organized as follows. The next section describes the general model. Section 3 presents the test for endogeneity and the estimation method. Section 4 describes results of a Monte-Carlo simulation. Section 5 presents an application to a scanner panel data set, and Section 6 concludes.

2. The Model

Consider the problem of consumers deciding whether to purchase in a certain product category, and if so, which brand to choose. Denote the consumer i's utility of a brand j in period t as

$$U_{ijt} = \gamma_{ijt} + \widetilde{X}_{ijt}\widetilde{\beta}_{it} + \mu_{ijt} \tag{1}$$

where $i \in \{1, \ldots, I\}$, $j \in \{0, \ldots, J\}$ with $j = 0$ representing no purchase, and $t \in \{1, \ldots, T\}$. The terms γ_{ijt} and μ_{ijt} represent an alternative specific effect for individual i in period t. These two terms are distinguished by different distribution assumptions as discussed below, with benefits in the estimation procedure (mixed logit model, McFadden and Train, 2000). The term γ_{ijt} captures the brand dummy in traditional logit models, allowing it to be different across individuals and through time, which will allow for common demand shocks and unobserved consumer heterogeneity. The term μ_{ijt} is the traditional logit error. The row vector $\widetilde{X}_{ijt}$ includes observable (to the researcher) brand characteristics relevant to the given consumer at the given time period, excluding brand dummies. Some of these variables may be constant across i (for example, several firms' marketing-mix variables such as price, feature, and display), or across t (consumer-specific characteristics). Suppose that there are k such variables so that the dimension of the vector $\widetilde{X}_{ijt}$ is k. The k-dimensional column vector $\widetilde{\beta}_{it}$ represents the weights given by consumer i in period t to the observable brand characteristics and is a vector of random coefficients.[9]

Equation (1) can be re-written in the following way. If we add to the vector $\widetilde{X}_{ijt}$ as additional elements the values of the brand dummies given that the observation is for alternative j (that is, all brand dummies take the value of zero except for the brand dummy of alternative j taking the value of one) we have a new vector of dimension $k + J$ which we denote as X_{ijt}. Adding in the same way to the vector $\widetilde{\beta}_{it}$ as additional elements the parameters associated with the brand dummies we have a new vector of dimension $k + J$ which we denote as β_{it}. Then equation (1) can be re-written as

$$U_{ijt} = X_{ijt}\beta_{it} + \mu_{ijt}. \tag{2}$$

We assume that

$$\beta_{it} = \beta_0 + \varepsilon_{it}, \tag{3}$$

where ε_{it} is distributed normally with mean 0 and variance-covariance matrix V_ε, and

$$\varepsilon_{it} = \psi_t + \eta_i + \xi_{it} \tag{4}$$

[9]This vector is not indexed by alternative j as no interaction is assumed between brand dummies and the other explanatory variables, as in typical choice models in the literature.

with independent mean-zero normal ψ_t, η_i, and ξ_{it}.[10] Note that $\varepsilon_{it}, \psi_t, \eta_i$, and ξ_{it} are $(k+J)$-dimensional column vectors. The term ψ_t represents the common shock in period t that is common across all consumers (note that it is a vector, potentially including a common shock per firm). Examples of such shocks include unobserved by the researcher or not explicitly modelled advertising, weather, public health announcements, competitor promotions, unaccounted for display features, availability or prices of other (not included in the model) products at the store, etc. Although ψ_t would not be observed by the researcher, firms could potentially observe it (at least partially), and make their marketing-mix decisions as a function of ψ_t (the potential endogeneity problem). The term η_i is specific to each consumer and common across all periods, and accounts for heterogeneity across consumers. Note that this specification accounts for additive and slope heterogeneity, as well as for common shocks. The individual and time specific term ξ_{it} accounts for random shocks to individual demands that are both time specific and not correlated across individuals, such as presence of guests at home or random taste changes, and could also be present due to an unobserved by the researcher targeted communication or promotion from the retailer to the consumer. The typical heterogeneity specification includes η_i but not ψ_t. On the other hand, Villas-Boas and Winer (1999) includes ψ_t but not η_i. Note that some of the variance is due to brand/period specific disturbances (i.e., it does not depend on i), and some of the variance is due to individual specific disturbances (i.e., may be uncorrelated across i), which may yield an arbitrary correlation pattern across i, j, and t. The error μ_{ijt} is assumed to be Gumbel distributed and independent of ε_{it}. The introduction of this error term allows us to have smooth probabilities of choice given ε_{it} and is known as the mixed logit model (McFadden and Train, 2000).[11] Note that identifying separately the variances of γ_{ijt} and μ_{ijt} is being done through the different distribution assumptions, which may lead to problems as discussed in Section 4 below (see also McFadden and Train). It is important to introduce γ_{ijt} to allow for the possibility of non-zero correlations across individuals for each period, and across periods for each individual.[12] The utility of the no purchase option is normalized to the error term $\mu_{i0,t}$. In period t consumer i chooses the alternative j which has the highest U_{ijt}.

The possibility of the common shock is a difference from the traditional models with independent and identically distributed across individuals error terms. Relaxing the independence across

---

[10] There may be serial correlation of ψ_t through time, as some unobservable variables (e.g., coupon drop) are likely to last for several time periods. If the process on ψ_t is such that the correlation between two time periods tends to zero when the time periods are further and further away, the estimators described below would still be consistent, by applying the results in Wooldridge (1994).

[11] McFadden and Train (2000) show that under certain regularity conditions, any random coefficient model can be approximated by a mixed logit model.

[12] Note that there are no constraints on the variances of the components of ε_{it} so that the variance of ξ_{it} can be set to zero without affecting any of the theoretical observations below.

individuals assumption on the error terms results in the possibility of endogeneity of the marketing-mix variables that are common across individuals, which may happen if (a part of) the common shock (which is not observed or modelled by the researcher) is known and taken into account by the retailer when setting (some of) the marketing variables X_{ijt}. This happens, for example, when the retailer jointly decides on the display and pricing options of products across categories, and the researcher only accounts for the display and pricing decisions in one category. Other examples would be competitive response of the retailer in the study to (unobserved to the econometrician) competitor's promotions, or the retailer's response to changes in weather or extra holiday demand (which is not explicitly accounted for in the econometric model). That is, we have to allow ψ_t to be correlated with some variables in X_{ijt}. In particular, in the application below, we consider a possible correlation with brand price, display, and feature. To estimate the parameters of the model, $\theta_0 = (\beta_0, V_\varepsilon)$, one then needs instruments for the variables in X_{ijt} that may be correlated with ψ_t. This issue is discussed in the next section.

Note that the specification above is sufficiently general to include both intercept and slope endogeneity, common shocks, and consumer heterogeneity. For example, for the case with no consumer heterogeneity, and only the possibility of common shocks and endogeneity on the "intercept" (the main case in Villas-Boas and Winer, 1999), equation (1) would reduce to

$$U_{ijt} = \beta_0^j + X_{ijt}^{-D}\beta_0^{-D} + \psi_t^j + \xi_{it}^j + \mu_{ijt} \tag{5}$$

where β_0^j, ψ_t^j, and ξ_{it}^j are the elements of the vectors β_0, ψ_t, and ξ_{it}, respectively, that multiply the dummy for brand j, where X_{ijt}^{-D} is the part of the vector X_{ijt} that does not have any brand dummies, and where β_0^{-D} is the part of the vector β_0 that is not associated with any brand dummies. For example, if the vector X_{ijt} includes brand dummies, price, display, and feature, then X_{ijt}^{-D} is a vector with the observations of price, display, and feature for brand j in period t, and β_0^{-D} is the vector of parameters associated with price, display, and feature. Note that both ξ_{it}^j and μ_{ijt} represent disturbances that are specific to the individual, time period, and alternative, the only difference being that ξ_{it}^j is normally distributed (and possibly correlated across j), and μ_{ijt} is Gumbel distributed. This specification is both more general and has computational advantages over the one with a single normally-distributed individual error term (see McFadden and Train, 2000).

3. Estimation

In order to estimate the parameters of the model, we propose a method of simulated moments that is roughly equivalent to the maximization of a quasi-likelihood function with and without instrumenting for the variables in X_{ijt}. In order to test for endogeneity we can compare an appropriately defined distance between the estimates (the parameters β_0 are identified up to a multiplicative constant) with a critical value of a χ^2 test statistic. The null hypothesis being tested is that ε_{it} is independent of X_{it}.

We can rewrite the model above in vector form as

$$U_{it} = X_{it}(\beta_0 + \varepsilon_{it}) + \mu_{it}, \tag{6}$$

where U_{it} is the column vector $(J \times 1)$ of utilities that consumer i assigns to the different alternatives at time t, X_{it} is the $(J \times (k + J))$ matrix of product and consumer characteristics relevant for consumer i at time t, β_0 is the parameter vector of mean weights of X_{it}, and ε_{it} is the vector of error terms. As stated above, ε_{it} are assumed identically normally distributed with mean zero and variance-covariance V_ε, and are pairwise independent when both i and t are different. Also as stated above, each element μ_{ijt} of the vector μ_{it} is Gumbel distributed, independent across i, j, and t, and independent of $\varepsilon_{i't'}$ for any i' and t'.

3.1. Estimation under the Null

Consider the estimation procedure under the null hypothesis that there is no endogeneity, that is, under the assumption that X_{it} is independent of ε_{it}.

Let the probabilistic event A_{ijt} represent the individual i choosing product j in period t. That is

$$A_{ijt} \equiv \{\text{purchase of } j^{\text{th}} \text{ brand by } i^{\text{th}} \text{ consumer in period } t\}.$$

Let also

$$d_{ijt} = 1_{A_{ijt}} \tag{7}$$

be the purchase indicator, i.e., $d_{ijt} = 1$ if and only if consumer i has bought brand j in period t, and $d_{ijt} = 0$ otherwise. Given the model assumptions, consumer i chooses alternative j at time t (where alternatives $j = 1, \ldots, J$ represent brand choice and alternative $j = 0$ represents no purchase) if and only if the utility of that alternative is at least as large as the utility of any other alternative at that time, i.e., $U_{ijt} \geq U_{ij't}$ for any $j' \neq j$. This condition can be rewritten as $B_j U_{it} \geq 0$, where $B_j = 1_j - I$ is a $J \times J$ matrix that is equal to the difference between the $J \times J$ matrix 1_j with 1's in the j^{th} column and zeros everywhere else, and the JxJ identity matrix I.

Hence, we have:

$$\begin{aligned}\Pr(d_{ijt}=1|X_{it}) &= \Pr(B_j U_{it} \geq 0\,|\,X_{it}) \\ &= \int \frac{e^{X_{ijt}(\beta_0+u)}}{1+\sum_{j'=1}^{J} e^{X_{ij't}(\beta_0+u)}} f_{V_\varepsilon}(u)\ du \equiv P_{ijt}(\theta_0),\end{aligned} \tag{8}$$

where $u \in R^J$, $\theta_0 = (\beta_0, V_\varepsilon)$, and $f_{V_\varepsilon}(u)$ is the density of the normal variable with mean vector 0 and variance-covariance matrix V_ε. The term $P_{ijt}(\theta_0)$ is the marginal probability of individual i choosing alternative j in period t. Multiplying these marginal probabilities evaluated at θ across i and t for the alternatives chosen by each individual i in every period t and taking the logarithm, one obtains

$$L_{null}(\theta) = \sum_{i,j,t} d_{ijt} \log P_{ijt}(\theta), \tag{9}$$

which is known as a quasi-loglikelihood function of the sample (see, e.g., Wooldridge, 1994). It is called a quasi-loglikelihood since it is the logarithm of the product of the marginal likelihoods of each observation and not the logarithm of the joint likelihood of the whole sample. If all the observations are independent (which would be the case without heterogeneity and without common shocks) this function would be the true loglikelihood of the sample. However, given that we allow for both heterogeneity and common shocks, the marginal probabilities of the different observations may not be independent across t (heterogeneity) or across i (common shocks), and therefore, the function above is a quasi-loglikelihood.

The estimator of θ that we propose is going to be based on the maximization of this quasi-likelihood function. In what follows, we construct the estimator given the complexity problems of calculation of $P_{ijt}(\theta)$ and show that, even though it is based on the quasi-likelihood function, it is consistent when both the number of households and the number of periods goes to infinity.

The numerical calculation of $P_{ijt}(\theta)$ with an appropriate precision is, for a sufficiently large number of alternatives, too time consuming. Therefore, we follow the method of simulated moments suggested in McFadden (1989). Namely, using $\sum_{i,j,t} P_{ijt}(\theta) = IT \equiv N$, and hence that $\sum_{i,j,t} \frac{\partial \log P_{ijt}(\theta)}{\partial \theta} = 0$, we can rewrite the first order conditions for maximization of $L_{null}(\theta)$ as

$$\sum_{i,j,t} (d_{ijt} - P_{ijt}(\theta)) \frac{\partial \log P_{ijt}(\theta)}{\partial \theta} = 0, \tag{10}$$

which can be solved by iteratively solving

$$\sum_{i,j,t}(d_{ijt} - P_{ijt}(\theta))\frac{\partial \log P_{ijt}(\theta_s)}{\partial\theta} = 0, \tag{11}$$

where θ_s is a consistent estimate of θ, possibly the solution of the previous iteration. The full estimation procedure is presented in the Appendix.

The maximization problem is thus rewritten as a method of moments in problem (11), in which $P_{ijt}(\theta)$ can be simulated as an average of individual predicted market shares resulting from random vector draws u from the distribution $f_{V_\varepsilon}()$.[13] Similarly, $\frac{\partial \log P_{ijt}(\theta_s)}{\partial\theta}$ can be simulated by taking the derivative of the logarithm of the probability $P_{ijt}(\theta_s)$ simulated with random draws independent from the ones used in the simulation of $P_{ijt}(\theta))$ above.[14] Let

$$g_{it}(\theta) = \sum_{j}(d_{ijt} - \widehat{P}_{ijt})\frac{\partial \log \widehat{P}_{ijt}(\theta_s)}{\partial\theta}, \tag{12}$$

where hats indicate the use of the simulated values instead of the precise probabilities. Let

$$g(\theta) = \frac{1}{N}\sum_{i,t} g_{it}(\theta). \tag{13}$$

Then the solution $\widehat{\theta}$ of the equation

$$g(\widehat{\theta}) = 0 \tag{14}$$

is our estimate of θ_0 under the null hypothesis. Note that since $\frac{\partial}{\partial\theta}$ of a function is a dim(θ)-dimensional vector, $g_{it}(\theta)$, and $g(\theta)$ are all dim(θ)-dimensional vectors, and thus the above equation is a system of dim(θ) equations on dim(θ) variables in θ. This system can be solved by minimizing $g(\theta)g(\theta)'$ over θ with standard numerical minimization techniques.

To estimate the standard deviation of the parameter estimates $\widehat{\theta}$ and prove consistency, we use

[13]Given each simulated draw of the normally distributed part of the error term, the probability reduces to the one given by the logit demand due to the Gumbel error term. These probabilities are then averaged across several draws of the normally-distributed error term to obtain the simulated probability $\widehat{P}_{ijt}(\theta)$.

[14]In other words, the simulation of the derivative of the logarithm of the probability can be done in exactly the same way as the simulation of the probability in $(d_{ijt} - P_{ijt}(\theta))$ part, but should be done with independent draws (McFadden 1989). Simulation of the probability in $(d_{ijt} - P_{ijt}(\theta))$ can be done with a small number of draws, since the errors of simulation will average out across different pairs of (i,t). The errors of the simulation of $\frac{\partial \log P_{ijt}(\theta_s)}{\partial\theta}$ will not average out since the derivative of the logarithm is not a linear function. However, it is not important to have these simulated very precisely, since the simulated $\frac{\partial \log \widehat{P}_{ijt}(\theta_s)}{\partial\theta}$ act as instruments, i.e., they only need to be correlated with the true $\frac{\partial \log \widehat{P}_{ijt}(\theta_s)}{\partial\theta}$ and do not have to be exactly equal (McFadden 1989).

a Taylor's expansion of $g(\theta)$ around θ_0:

$$g(\widehat{\theta}) = g(\theta_0) + G(\overline{\theta})(\widehat{\theta} - \theta_0) = 0, \tag{15}$$

where $\overline{\theta}$ is on the interval connecting θ_0 and $\widehat{\theta}$ and $G(\overline{\theta}) = \frac{\partial g(\overline{\theta})}{\partial \theta}$. We can write $\widehat{\theta} - \theta_0$ as

$$\widehat{\theta} - \theta_0 = -G(\overline{\theta})^{-1} g(\theta_0). \tag{16}$$

Note that because $E(d_{ijt} - P_{ijt}(\theta))$ is equal to zero at $\theta = \theta_0$, we have $Eg(\theta_0) = 0$, and then it follows from (16) that $E(\widehat{\theta} - \theta_0) = 0$.

Let $K = \min\{I, T\}$. Since we assumed that ε_{it} and $\varepsilon_{i't'}$ are independent unless $i = i'$ or $t = t'$ (and hence so are d_{ijt} and $d_{i'j't'}$),

$$\text{Var}(g) = \frac{1}{N^2} \sum_{i,t,i',t'} \text{Cov}(g_{it}, g_{i't'}) = \frac{1}{N^2} \sum_{(i,t,i',t') \in C} \text{Cov}(g_{it}, g_{i't'}) = O\left(\frac{1}{K}\right), \tag{17}$$

where $C = \{(i,t,i',t') \mid i = i' \text{ or } t = t'\}$ and $O\left(\frac{1}{K}\right)$ represents a sequence that goes to zero at the rate of at least $\frac{1}{K}$, i.e., $\limsup_{k \to \infty} KO(\frac{1}{K})$ is finite. Therefore,

$$\text{Var}(\widehat{\theta}) \to 0 \text{ as } K \to \infty, \tag{18}$$

and hence, the estimator $\widehat{\theta}$ is consistent as long as we make both the number of periods and the number of consumers go to infinity. The number of periods T has to go to infinity in order to average out the common shocks ψ_t. The number of consumers I has to go to infinity in order to average out the unobserved heterogeneity η_i. Then by the central limit theorem when $K \to \infty$,

$$\sqrt{K}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, G^{-1} K \text{Var}(g) G^{-1'}). \tag{19}$$

Furthermore, $\text{Var}(g)$ can be approximated by

$$\widehat{V}_{gg} \equiv \widehat{\text{Var}}(g) = \frac{1}{N^2} \sum_{(i,t,i',t') \in C} g_{it}(\widehat{\theta}) g'_{i't'}(\widehat{\theta}). \tag{20}$$

This variance estimator has an asymptotic error of order $\frac{1}{K}$ as well, and hence is also consistent. Note, that the above formula includes the additional variance introduced by the simulation of $P_{ijt}(\theta)$ (see McFadden 1989).

Note that for the consistency of the estimator described above, it is not crucial that we observe all T periods of any given household. In fact, if each household that is introduced in the sample is only there for M periods, for any $M \geq 1$, the consistency results above follow in exactly the same way. This point can be important if one wants to apply the results above to purchase models conditional on choice, where the number of purchases per household in the typical sample is less than 10 and often only 3 or 4 : even if the number of purchases per household is small, as far as the number of households that make purchases is large enough, and the purchases are made in large enough number of periods, one can use the above method. A small number of purchases per individual (keeping the number of purchases constant) would mean that the correlation of error terms within an individual may be less of a problem (as, relative to the data with large number of purchases per individual, such data would have more observations with independent error terms). Similarly, the consistency results also follow if we observe a small number of purchases of each brand per week (or no purchases in some weeks).

3.2. Estimation of the Mean Parameters with Instrumental Variables.

Consider now that some columns of X_{it} may be correlated with ε_{it}. Namely, let the indices of the columns of X_{it} that are correlated with ε_{it} be from $M \subset \{1, \ldots, k\}$, where k is the number of columns in X_{it} (dimension of β). In order to simplify the analysis (but not a crucial assumption, see below) suppose that the only elements of ε_{it} that have non-zero variance are the ones associated with the brand dummies, so that we have $X_{it}\beta_{it} = X_{it}\beta_0 + \varepsilon_{it}$. In this setting, we can then denote by ε_{it} the vector of dimension J of error terms that interact with brand dummies. Denote the m^{th} column of X_{it} by $X_{it}^{(m)}$, and suppose we have instruments $Z_{it}^{(m)}$ for $X_{it}^{(m)}$, where $m \in M$, i.e., random variables which are correlated with $X_{it}^{(m)}$ and are independent of ε_{it}.[15] Then we can estimate β_0 (up to a multiplicative constant) through the following two-step procedure.

As we only worry about endogeneity of the variables in X_{it} that do not actually depend on i, we omit the subscript i from those variables (as well as from the instruments Z_{it}; we assume that the instruments are the same for all i). Assume that the dependence of $X_t^{(m)}$ on $Z_t^{(m)}$ is linear:

$$X_t^{(m)} = Z_t^{(m)\prime}\pi_0^{(m)} + \zeta_t^{(m)}, \tag{21}$$

[15] The possibility of the instruments being independent of the error terms could potentially be tested with a specification-type test comparing estimates with different sets of possibly valid instruments (Hausman 1978). Tests for the possibility of the instruments not being correlated with the endogenous right hand side variables (weak instruments problem) can be seen in Hahn and Hausman (2002), and references cited there. This possibility would potentially lead to high standard errors, and be reflected in the estimates.

where $\zeta_t^{(m)}$ are independent normal random variables, identically distributed for each m.[16] Regressing $X_t^{(m)}$ on $Z_t^{(m)}$, one obtains a consistent estimator $\widehat{\pi}^{(m)}$ of $\pi_0^{(m)}$. Let $\widehat{X}_{it}$ be the matrix obtained from X_{it} by replacing each column with index from M by the corresponding $Z_t^{(m)'}\widehat{\pi}^{(m)}$ (repeated for each i). For $X_{it}^{(m)}$ with no endogeneity, we just have $\widehat{X}_{it}^{(m)} = X_{it}^{(m)}$. The computation of $\widehat{X}_{it}$ is the first stage of the instrumental variables estimation.

For the second stage, rewrite U_{it} as follows:

$$U_{it} = X_{it}\beta_0 + \varepsilon_{it} + \mu_{it} = \widehat{X}_{it}\beta_0 + \zeta_t\beta_0 + \varepsilon_{it} + \mu_{it} = \widehat{X}_{it}\beta_0 + \widetilde{\varepsilon}_{it} + \mu_{it}, \tag{22}$$

where $X_{it} = \widehat{X}_{it} + \zeta_t$, and $\widetilde{\varepsilon}_{it} = \zeta_t\beta_0 + \varepsilon_{it}$.

We have that $\widetilde{\varepsilon}_{it}$ are normal random variables, identically distributed across i and t, and independent of $\widetilde{\varepsilon}_{i't'}$ unless $i = i'$ or $t = t'$. Hence we can apply the analysis of the previous subsection to the model

$$U_{it} = \widehat{X}_{it}\beta_0 + \widetilde{\varepsilon}_{it} + \mu_{it}, \tag{23}$$

once we know $\widehat{X}_{it}$.[17] Let us denote the parameters of this instrumental variables estimation by $\tau = (\beta, V_{\widetilde{\varepsilon}})$ where $V_{\widetilde{\varepsilon}}$ is the variance matrix of $\widetilde{\varepsilon}_{it}$. Note that because we only have choice data, β_0 is only estimated up to a constant. This simulated GMM estimation is the second stage.[18]

We now turn to the question of estimation of the standard errors and consistency of these parameter estimators. The previous subsection gives us the standard errors for estimation of (23) in the case where the values of $\pi_0^{(m)}$ are known. In our case, however, we know them with some error (from the OLS regression in the first stage), and therefore we have to adjust the standard errors to account for the bias coming from the first stage (errors of $\widehat{\pi}$ estimates). Since $\widehat{\pi} \xrightarrow{\text{p}} \pi_0$, the estimate is again consistent and asymptotically normal (by the same proof as above). The formula and derivation of an estimator for the variance matrix of $\widehat{\tau} - \tau_0$ (where $\tau_0 = (\beta_0, V_{\widetilde{\varepsilon}})$) is presented in the Appendix. In the Appendix we also present an estimator for the variance V_ε of the error terms ε_{it}.

[16] Serial correlation on ζ_t would not affect the consistency of the estimator but only its asymptotic variance as long as the correlation goes to zero with time.

[17] Alternatively, one could consider as explanatory variables in (23) not only $\widehat{X}_{it}$ but the estimated ζ_t as well, where the parameters associated with ζ_t includes both β_0 and the correlations of $\zeta_t\beta_0$ with ε_{it}. This would involve additional $J \times J$ parameters to be estimated, if random coefficients only on the brand dummies as described here. The Appendix presents further details on the estimation of these correlations, and how to use them to compute V_ε. Petrin and Train (2002) present this approach of including the estimated ζ_t as explanatory variables in (23).

[18] The case where ε_{it} also interacts with time-varying X_{it} can be dealt with in the set-up above, but now the error term $\widetilde{\varepsilon}_{it}$ of the adjusted latent utility (22) will have a distribution that is also a function of the fitted X_{it}. This case has been known as the "slope endogeneity" case. See Section 6 in Villas-Boas and Winer (1994) for an analysis of that case under maximum likelihood. See also Manchanda et al. (2004) and Luan and Sudhir (2006).

Note that by using the estimates of V_ε and β_0 one can construct estimators of the variance of the consumer heterogeneity, V_η, of the demand common shocks, V_ψ, and of the consumer-period errors, V_ξ, by using fixed effects for periods and consumers, possibly in an iterative procedure. We would then be able to decompose the variance of the errors into heterogeneity, common shocks, and consumer-period specific errors, i.e., $V_\varepsilon = V_\eta + V_\psi + V_\xi$. It would be interesting to learn the small sample properties of the estimation of this decomposition of the variance of the errors.

Finally, note that the specification above can be made consistent with a structural model of firm decisions for appropriate assumptions on the firms' cost functions, if the endogenous variable structural equations can be inverted as a function of the firms' costs. After inverting the endogenous variable structural equations, one can then immediately obtain the conditions on the firms' cost functions such that the model above is consistent with the structural model considered.

3.3. A Test of Endogeneity

In order to test the endogeneity of some variables in X_{it} we may check whether the parameter estimates of β_0 with and without instrumenting for some of the variables in X_{it} are statistically different under the null hypothesis that X_{it} and ε_{it} are independent up to a multiplicative constant.

Let

$$h_{it}(\widehat{\pi}, \tau) = \sum_j (d_{ijt} - \widehat{P}_{ijt}(\tau \mid \widehat{X}_{it})) \frac{\partial \log \widehat{P}_{ijt}(\tau_s \mid \widehat{X}_{it})}{\partial \tau}, \tag{24}$$

where τ_s is again a fixed estimate of τ, $\widehat{\pi}$ is the value of π used in $\widehat{X}_{it}$ that estimates X_{it}, and the draws used for the simulation of $\widehat{P}_{ijt}(\tau \mid \widehat{X}_{it})$ and $\frac{\partial \log \widehat{P}_{ijt}(\tau_s \mid \widehat{X}_{it})}{\partial \tau}$ are independent. Also, let $h(\widehat{\pi}, \tau) = \sum_{i,t} h_{it}(\widehat{\pi}, \tau)$, and $H(\widehat{\pi}, \tau) = \frac{\partial h(\widehat{\pi}, \tau)}{\partial \tau}$.

Then, similarly to (16), we obtain

$$\widehat{\tau} - \tau_0 = -H^{-1}(\widehat{\pi}, \overline{\tau}) \sum_{i,t} h_{it}(\widehat{\pi}, \tau_0), \tag{25}$$

for some $\overline{\tau}$ on the interval connecting $\widehat{\tau}$ and τ_0. This means that under the null hypothesis, the asymptotic variance of $\widehat{\tau}$ is

$$H^{-1}(\widehat{\pi}, \overline{\tau}) \left(\sum_{(i,t,i',t') \in C} \mathrm{Cov}(h_{it}(\widehat{\pi}, \tau_0), h_{i't'}(\widehat{\pi}, \tau_0)) \right) H^{'-1}(\widehat{\pi}, \overline{\tau}). \tag{26}$$

This asymptotic variance can be consistently estimated as

$$\widehat{\text{Var}}(\widehat{\tau}-\tau_0)=H^{-1}(\widehat{\pi},\widehat{\tau})\left(\sum_{(i,t,i',t')\in C}\widehat{h}_{it}\widehat{h}'_{i't'}\right)H'^{-1}(\widehat{\pi},\widehat{\tau}), \tag{27}$$

where $\widehat{h}_{it}$ is h_{it} evaluated at $(\widehat{\pi},\widehat{\tau})$. This expression is obtained because evaluating h at $\widehat{\pi}$ does not bring additional covariances as $\widehat{\pi}$ is a function of X_{it} and Z_t which are both independent of ε_{it} (by the null hypothesis and by assumption).

Denote the two β_0 estimates obtained through $\widehat{\theta}$ and $\widehat{\tau}$ as $\widehat{\beta}_1$ and $\widehat{\beta}_2$, respectively. Then, using (16) and (25), we have that for some scalar c,

$$\widehat{\beta}_1-c\widehat{\beta}_2=\sum_{i,t}\left(cH^{-1}(\overline{\pi},\overline{\tau})h_{it}(\widehat{\pi},\tau_0)-G^{-1}(\overline{\theta})g_{it}\right)_{\beta}, \tag{28}$$

where the subscript β indicates that the vector has to be truncated to be of the same length as β. Hence, for that c, $\widehat{\beta}_1-c\widehat{\beta}_2$ is asymptotically normally distributed with mean zero, and variance given by

$$\widehat{V}_{test}=\left(\widehat{H}^{-1}\widehat{V}_{hh}\widehat{H}'^{-1}+c^2\widehat{G}^{-1}\widehat{V}_{gg}\widehat{G}'^{-1}-c(\widehat{H}^{-1}\widehat{V}_{hg}\widehat{G}'^{-1}+\widehat{G}^{-1}\widehat{V}_{gh}\widehat{H}'^{-1})\right)_{\beta\beta}, \tag{29}$$

with $\widehat{H}=H(\widehat{\pi},\widehat{\tau})$, and $\widehat{G}=G(\widehat{\theta})$, and where

$$\widehat{V}_{ab}=\sum_{(i,t,i',t')\in C}a_{it}b'_{i't'}, \tag{30}$$

for each a and b, the symbolic subscripts used in the previous equation, and the subscript $\beta\beta$ stands for taking the first kxk submatrix, where k is the dimension of β. The test then should compare the value of

$$\text{test}=\text{test}(c)=(\widehat{\beta}_1-c\widehat{\beta}_2)'\widehat{V}_{test}^{-1}(\widehat{\beta}_1-c\widehat{\beta}_2) \tag{31}$$

with the critical value of $\chi^2_{(k)}$. The value of c is unknown, but we can use $c=\frac{\widehat{\beta}_1^{(i)}}{\widehat{\beta}_2^{(i)}}$ for some i. In the application below we use the price coefficient for i, and compare if the ratio of the other parameters to the price coefficient changes when we use instrumental variables.[19] Note that as both the estimation under the null and with instrumental variables involves some loss of efficiency (quasi-likelihood function and method of simulated moments), the parameters are estimated with

[19] Alternatively, one can minimize the Wald criterion over c as noted in Hausman and Ruud (1987). This results in a greater difference between the estimates with and without accounting for endogeneity in the application below. We thank Sanjog Misra for this point.

less precision, and therefore this test can be seen as a more conservative test of endogeneity (more likely not to reject exogeneity even if there actually is endogeneity).

4. Monte-Carlo Simulation

One potential concern with the estimation method presented above is whether the number of households or time periods that is typically available is sufficient for the asymptotic results to work. In order to have further information on this question we have performed the following Monte-Carlo study.

We simulate a 6-brand purchase data set with 112 time periods and 300 households (these numbers of time periods and households are typically found in the available scanner panel data sets). We set the brand dummies' parameters to be 3, the price coefficient to be -4, and the display and feature coefficients to be 2. In the simulated household utility error term, we introduce heterogeneity and common shocks terms, with only the error terms associated with the brand dummies having non-zero variance (common shocks and heterogeneity in the intercept) as considered above and in the application of the next Section. The variance of ε_{ijt} was set at one, with zero correlation across j, i.e., V_ε is the identity matrix. The variances of ψ_t, η_i, and ξ_{it} were all set at 1/3 times the identity matrix.[20] We also simulate the firm response variables, in which we assume that price, display, and feature depend both on the (simulated) instruments and on the utility common shock.

The simulated instruments are vectors Z_1, Z_2, and Z_3, each consisting of twenty time-varying variables each, all independent, normally distributed, and with variance .03125. The variables in Z_1 have mean 1.5, while those in Z_2 and Z_3 have mean zero. Price for brand j in period t was generated as $P_{jt} = \sum_{k=1}^{20} Z_{1kt}/\sqrt{20} + \zeta_{jt}$ where ζ_{jt} is normally distributed with mean zero and variance .03125, has correlation one with ψ_{jt}, and is independent of all other errors.[21] Display for brand j in period t is either one or zero, taking the value of one if and only if $\sum_{k=1}^{20} Z_{2kt}/\sqrt{20} + \zeta_{jt} > 0$. Finally, the

[20] Allowing the common shock and heterogeneity to be as high as equally splitting the error term (making the variance of individual-week specific error ξ_{it} equal to zero) resulted in only slightly increased estimation error relative to the ones reported below.

[21] The reason for the instrumental variables number and variance selection is as follows. First, we would like to have prices roughly between 1 and 2 (to be in-line with the data we use in the following section). We do that by making the price a normal random variable with mean 1.5 and variance 0.0625 (so that 1.5 plus minus the double of its standard deviation would be the interval 1 to 2). Second, to be roughly consistent with the R^2 of the price regression on instruments in the application of the following section, we would like the instruments to explain half of the variance, thus assigning variances of $\sum_{k=1}^{20} Z_{1kt}/\sqrt{20}$ and ζ_{jt} to be the same. This yields the variance .03125 for each of Z_{1kt} and ζ_{jt}. Note that the reason for using 20 instrumental variables is again to be consistent with the number of instruments used in the application of the following section. Similarly generating instruments for display and feature results in a lower value of R^2 because of the conversion of these variables to dummy variables.

Parameter	Null	Null Normalized	Instr	Instr adj	Simulated
brand 1	0.724 (0.191)	1.019 (0.256)	4.748 (7.409)	2.762 (0.515)	3.0
brand 2	0.678 (0.209)	0.956 (0.288)	4.307 (4.512)	2.695 (0.553)	3.0
brand 3	0.709 (0.202)	0.999 (0.273)	4.886 (7.869)	2.828 (0.465)	3.0
brand 4	0.728 (0.188)	1.026 (0.257)	4.837 (7.413)	2.827 (0.537)	3.0
brand 5	0.699 (0.190)	0.984 (0.258)	4.886 (8.460)	2.749 (0.590)	3.0
brand 6	0.718 (0.215)	1.012 (0.295)	4.733 (6.541)	2.803 (0.550)	3.0
price	-2.83 (0.169)	-4.0 (0.000)	-7.06 (11.77)	-4.0 (0.0)	-4.0
display	2.391 (0.125)	3.373 (0.124)	3.407 (5.333)	1.981 (0.265)	2.0
feature	2.400 (0.134)	3.384 (0.115)	3.529 (5.826)	2.012 (0.215)	2.0

Table 1: Results of the Monte Carlo study. The standard errors across simulations are in parentheses. "Null" means without instrumenting; "instr" means when instrumenting for price, display, and feature; and "instr adj" means the instrumental variables estimation with rescaling of the parameters so that the price coefficient would equal to true one for each simulation. "Simulated" is the true parameter values in the simulation.

variable feature for brand j in period t is again either one or zero, taking the value of one if and only if $\sum_{k=1}^{20} Z_{3kt}/\sqrt{20}+\zeta_{jt} > 0$. The resulting R^2 statistic of the regression of the price on instruments is about .5, and of the dummy display and feature variables on the instruments is about .4. Further, the estimation is performed with and without instrumenting for price, display, and feature. The simulation and estimations are repeated 100 times to generate sufficient information for the empirical means and standard deviations of the parameter estimates.

The results are presented in Table 1. One can see that when price is correlated with the utility common shocks, the estimation not accounting for this correlation yields biased results (columns 1 and 2 as compared to the last column). However, using instrumental variables in the estimation corrects for this problem if the parameter estimates are considered up to a scalar. The larger values and large variances in the instrumental variables estimation without normalizing by one of the parameters are due to a few instances when parameter estimates tended to be very large. This indicates that the decomposition of the error term into the normal and Gumbel component can not be well recovered. However, if the estimates are normalized by one of the mean parameters, or by the variance of the utility error term (and not just by its logit component), then the convergence seems relatively good. After normalization by the price estimate, the brand intercepts are still a little smaller than they should be. This is probably due to the small sample fitting of the instruments to price.[22] Table 1 also indicates that the fact that display and feature were simulated to be dummy variables and not normal, did not affect the results too much in this simulation. We should state

[22]The brand intercepts, along with other parameters, were on average correctly recovered when we used one instrument (again, explaining half of the variance of price) instead of twenty, for each price, display, and feature.

Brand	Price	Display	Feature	Market share
DelMonte 32 Oz	\$1.17	.30	.083	.24
Heinz 32 Oz	\$1.29	.20	.068	.21
Hunts 32 Oz	\$1.21	.13	.053	.13
Heinz 44 Oz	\$1.85	.09	.023	.07
Brook 32 Oz	\$.99	.59	.034	.06
Heinz 28 Oz	\$1.31	.19	.032	.06

Table 2: Sample means.

again that these small sample results depend on a limited model simulation, and, therefore, their generality should be further tested.

5. Application to a Scanner Panel Data Set

5.1. Data and Complete Model Specification

We estimated the model on scanner panel data of a ketchup market in the mid-80's. We restricted attention to one store where we had a sample of 340 households over a period of 132 weeks, with a total of 1445 purchases. We restricted attention the six brands with the greater market share, which accounted for 77% of the market. These brands are DelMonte 32 Oz, Heinz 32 Oz, Hunts 32 Oz, Heinz 44 Oz, Brook 32 Oz, and Heinz 28 Oz. Table 2 presents summary statistics for these six brands.

The vector of explanatory variables in the latent utility model, X_{ijt}, is composed of brand dummies, price, display dummy, feature dummy, and the brand loyalty variable.

The brand loyalty variable of a household is calculated using the first 20 weeks in the data as the number of times that brand was purchased by the household in the first 20 weeks. The loyalty variable, thus, does not change with t. The loyalty variable is included in the tradition of the literature on brand choice models (e.g., Guadagni and Little 1983).[23] Another potential issue, not dealt with in this paper, is if there is a time pattern to price dynamics with some consumers refraining from purchase in anticipation of price discounts, and purchasing then in large quantities to stockpile for future consumption occasions. For some discussion of these issues see, for example, Blattberg et al. (1981), Jeuland and Narasimhan (1985), Gupta (1988), van Heerde et al. (2003), Sun et al. (2003), Macé and Neslin (2004). These lagged effects can generate important dynamic

[23]Note that the loyalty variable may be correlated with the household heterogeneity error term, and therefore bring bias into the estimation procedure. This problem also applies in general to the heterogeneity literature that includes loyalty variables. See Honoré and Kyriazidou (2000) and Chintagunta et al. (2000) for a discussion of this problem and possible solutions.

implications, which can seriously complicate the estimation procedure. Sun et al. (2003) present simpler reduced form variables to deal with these issues. If the consumption rate is sufficiently high in comparison to the frequency of purchase, or the consumption rate or inventory costs are sufficiently stochastic, one would expect that the inventory effects would be smaller, and therefore, not accounting for these effects would have less impact on the parameter estimates. It would be interesting to investigate how these effects may interact with accounting for endogeneity, but this is beyond the scope of this paper.

The parameters to be estimated are the ones in the vector β_0 with dimension nine (six brand dummies, price, display, and feature) plus the variance matrix V_ε.

In the estimation we are worried about potential endogeneity of price, display, and feature. As a vector of instruments, we used the interaction between the brand dummies and the prices of different types of tomatoes in the current period and with lags of five, ten, fifteen, and twenty weeks.[24] Different lags did not affect the results. The average R^2 for the regression of prices on these instruments was .41. The R^2 for the regression of display and feature on these instruments was between .2 and .3.[25] Although one might not think of a direct effect of these instruments on the display and feature decisions, these instruments may be correlated with the display and feature decisions, because they may affect prices, and prices interact in the profit function with the display and feature decisions. Finding other instruments that are more directly correlated with the display and feature decisions is an important issue for future research. The application presented uses data only from purchases from one store. Using data from purchase in multiple stores is possible (by augmenting the number of alternatives), but it requires more computational capability.

5.2. Results and Discussion

The results of the estimation of the method of simulated moments described above are reported in Table 3. The left and middle columns present the results without ("null") and with ("instr") instrumenting for price, display, and feature, respectively. The parameters are normalized by the variance of the extreme-value part of the error term. Note that because the error term may be different in the two cases, the scaling of the latent utility may also be different. However, regardless of

[24] These price series were obtained from the United States Department of Agriculture.

[25] Given that display and feature are dummy variables, the error terms of the regression of these variables on the instruments cannot be normally distributed as assumed above. This can be a potential issue in the estimation. In the Monte-Carlo study described in Section 4, we have seen that this misspecification has not caused a serious bias in the parameter estimates. We have also run the model with labor rates and energy prices as instruments and the main message of the results did not change.

Parameter	Null	Instr	Instr adj
DelMonte 32 Oz	-2.989 (0.884)	-2.034 (1.188)	-1.508 (0.881)
Heinz 32 Oz	-4.228 (2.295)	-2.036 (2.459)	-1.510 (1.824)
Hunts 32 Oz	-4.353 (3.265)	-5.429 (3.458)	-4.027 (2.565)
Heinz 44 Oz	-3.672 (1.161)	-2.154 (1.386)	-1.597 (1.028)
Brook 32 Oz	-4.114 (3.582)	-3.289 (4.408)	-2.439 (3.269)
Heinz 28 Oz	-2.882 (1.704)	-1.906 (1.543)	-1.413 (1.144)
price	-1.824 (0.560)	-2.459 (0.705)	-1.824 (0.523)
display	0.238 (0.442)	-0.114 (1.190)	-0.084 (0.882)
feature	1.664 (0.216)	1.471 (0.405)	1.091 (0.300)
loyalty	0.794 (0.198)	0.735 (0.197)	0.545 (0.146)

Table 3: Results for market data. The standard errors are in parentheses. "Null" means without instrumenting; "instr" means when instrumenting for price, display, and feature; and "instr adj" means the instrumental variables estimation with rescaling the price coefficient by the "null" results.

the scaling factor, the estimation suggests that without accounting for endogeneity, the importance of price relative to feature is underestimated.

In order to compare the parameters across the two estimations, the right column ("instr adj") represents the instrumental variables estimates if the price is normalized so as to make it equal to the price parameter of the estimation without instrumenting. The standard errors of the parameters were obtained through bootstrap. This incorporates into the standard errors the heterogeneity, common shocks, two-step estimation, and simulations issues.

Comparing the left and right columns, one can see that the relative impact of price and feature with respect to the other explanatory variables increases on an average of about 60% when one instruments for the potential endogenous variables. That is, endogeneity in prices causes under-estimation of its importance with respect to the other variables. Note also that this estimation takes into account possible consumer heterogeneity on the brand dummies. Interestingly, it seems that endogeneity does not seem to be a serious problem in this data set on the display or feature variables.

Performing the statistical test described above to the question of whether the relative value of the mean parameters are the same with and without instrumenting, one obtains the value of the statistic equal to 36.4. Since the statistic is distributed χ^2 with nine degrees of freedom, the relative value of the mean parameters is indeed statistically different at the 5% significance level.

The estimation procedure described above argues that for a sufficiently large number of both households and time periods in the sample, consumer heterogeneity or common shocks do not cause inconsistency in the traditional estimators (e.g., Guadagni and Little, 1983), although they affect the standard errors. However, endogeneity (correlation between the error terms and the explanatory variables) may cause the estimators to be inconsistent.

In the consumer heterogeneity literature (e.g., Kamakura and Russell 1989) one often finds that the estimates of the parameters change when one accounts for consumer heterogeneity. Given the discussion above there are two potential explanations for this general finding. One possibility is that in fact the sample size is small and the asymptotics do not work, such that fully specifying the consumer heterogeneity into the likelihood function (or the moments condition) yields different results. Alternatively, it may be that because those heterogeneity estimates do not account for the potential endogeneity of the firms' decision variables, they are the result of a misspecified model. If such explanation were true then including heterogeneity in the model would just change the model misspecification without truly estimating consumer heterogeneity.

Which of these two explanations is actually true in the typical scanner panel data sets is an important research question, that ultimately depends on whether the sample size is large enough for the asymptotics to work.[26]

Note also that for the estimates from the above moment conditions to be close to the true parameters, we need the number of periods to be large enough. This does not require the number of purchases or observations per household to be large, since the common shock errors average out across consumers as long as they make purchases in different periods.

6. Concluding Remarks

This paper presents a method of consistent estimation of the parameters in a latent utility model when one has both consumer heterogeneity and endogeneity in a general error structure. The method is one of simulated moments where one averages out the heterogeneity terms and the common shocks across observations, but includes the added correlation among observations in the estimates of the standard errors. This method seems to be practical to implement with any number of alternatives, and allows for testing for endogeneity for several of the firms' actions.

In an application to a ketchup data set we tested for endogeneity in price, display, and feature, and found endogeneity problems in price but not necessarily in display or feature.

[26]For consumer heterogeneity the question is whether the number of households in the sample is large enough.

APPENDIX

1. *Derivation of an Estimator of the Variance of* $\widehat{\tau} - \tau_0$.

To account for the contribution of the asymptotic variance of $\widehat{\pi}$ to the asymptotic variance of the parameter estimates in (23), we follow the influence function approach suggested in Newey and McFadden (1994, p. 2178), for which we note that in the OLS regression,

$$\widehat{\pi} - \pi_0 = \sum_t (\frac{1}{T} Z'Z)^{-1} \frac{1}{T} Z_t'(X_t - Z_t\pi_0) = \sum_t D_{zz}^{-1} \frac{1}{T} Z_t'(X_t - Z_t\pi_0) + O\left(\frac{1}{T}\right) = \sum_t \lambda_t + O\left(\frac{1}{T}\right), \quad \text{(i)}$$

where $Z'Z = \sum_t Z_t'Z_t$, $D_{zz} = \text{plim}_{T\to\infty} \frac{1}{T} Z'Z$, and $\lambda_t = D_{zz}^{-1} \frac{1}{T} Z_t'(X_t - Z_t\pi_0)$ are independent across t. Let

$$h_{it}(\widehat{\pi}, \tau) = \sum_j (d_{ijt} - \widehat{P}_{ijt}(\tau \mid \widehat{X}_{it})) \frac{\partial \log \widehat{P}_{ijt}(\tau_s \mid \widehat{X}_{it})}{\partial \tau}, \quad \text{(ii)}$$

where τ_s is again a fixed estimate of τ, $\widehat{\pi}$ is the value of π used in $\widehat{X}_{it}$ that estimates X_{it}, and draws used for the simulation of $\widehat{P}_{ijt}(\tau \mid \widehat{X}_{it})$ and $\frac{\partial \log \widehat{P}_{ijt}(\tau_s \mid \widehat{X}_{it})}{\partial \tau}$ are independent. Also, let $h(\widehat{\pi}, \tau) = \sum_{i,t} h_{it}(\widehat{\pi}, \tau)$, $H(\widehat{\pi}, \tau) = \frac{\partial h(\widehat{\pi}, \tau)}{\partial \tau}$. Then our estimate $\widehat{\tau}$ of τ_0 is the solution of $h(\widehat{\pi}, \widehat{\tau}) = 0$.

In order to estimate the variance matrix of $\widehat{\tau} - \tau_0$ (where $\tau_0 = (\beta_0, V_{\widetilde{\varepsilon}})$), note that

$$\widehat{\tau} - \tau_0 = -H^{-1}(\widehat{\pi}, \overline{\tau}) h(\widehat{\pi}, \tau_0) = -H^{-1}(\widehat{\pi}, \overline{\tau}) \left(h(\pi_0, \tau_0) + H_\pi(\overline{\pi}, \tau_0)(\widehat{\pi} - \pi_0)\right), \quad \text{(iii)}$$

where $H_\pi(\pi, \tau) = \frac{\partial h(\pi, \tau)}{\partial \pi}$ and $\overline{\tau}$ and $\overline{\pi}$ are on the interval connecting $\widehat{\tau}$ and τ_0, and $\widehat{\pi}$ and π_0, respectively. Using this with the representation of $\widehat{\pi} - \pi_0$ in (i) we obtain the asymtotic equality

$$\widehat{\tau} - \tau_0 = -H^{-1}(\widehat{\pi}, \overline{\tau}) \sum_t \left(H_\pi(\overline{\pi}, \tau_0)\lambda_t + \sum_i h_{it}\right). \quad \text{(iv)}$$

It then follows that the asymptotic variance of $\widehat{\tau} - \tau_0$ is

$$\begin{aligned} \text{Var}(\widehat{\tau} - \tau_0) &\overset{\text{asy}}{=} H^{-1}(\widehat{\pi}, \overline{\tau}) \text{Var}\left(\sum_t \left(H_\pi(\overline{\pi}, \tau_0)\lambda_t + \sum_i h_{it}\right)\right) H'^{-1}(\widehat{\pi}, \overline{\tau}) \\ &\overset{\text{asy}}{=} H^{-1}(\widehat{\pi}, \overline{\tau}) \left(\sum_{i,t,t',t\neq t'} \text{Cov}(h_{it}, h_{it'}) + \sum_t \text{Var}(h_t + H_\pi(\overline{\pi}, \tau_0)\lambda_t)\right) H'^{-1}(\widehat{\pi}, \overline{\tau}), \end{aligned}$$

where $h_t = \sum_i h_{it}$. Hence, the asymptotic variance of $\widehat{\tau} - \tau_0$ can be consistently estimated by (see

Newey and McFadden, 1994, p. 2183, equation (6.11))

$$H^{-1}(\widehat{\pi},\widehat{\tau})\left(\sum_{i,t,t',t\neq t'}\widehat{h}_{it}\widehat{h}'_{it'}+\sum_{t}(\widehat{h}_t+H_\pi(\widehat{\pi},\widehat{\tau})\widehat{\lambda}_t)(\widehat{h}_t+H_\pi(\widehat{\pi},\widehat{\tau})\widehat{\lambda}_t)'\right)H'^{-1}(\widehat{\pi},\widehat{\tau}), \tag{v}$$

where $\widehat{h}$ is h evaluated at $(\widehat{\pi},\widehat{\tau})$, and $\widehat{\lambda}_t=(Z'Z)^{-1}Z_t'(X_t-Z_t\widehat{\pi})$.

2. *Derivation of an Estimator of V_ε with Instrumental Variables.*

We obtain here an estimator for the variance of the error terms ε_{it}, V_ε. Note that in the instrumental variables estimation we obtain an estimator of the errors $\widetilde{\varepsilon}_{it}$, $V_{\widetilde{\varepsilon}}$. Because of the assumed joint normality of ζ_t and ε_{it} we can write

$$\varepsilon_{it}=\rho(\zeta_t\beta_0)+\nu_{it}, \tag{vi}$$

where ρ is a JxJ matrix and ν_{it} is independent of ζ_t. Then

$$\Pr(d_{ijt}=1|X_{it})=\Pr(B_j\nu_{it}\geq -B_j(X_{it}\beta_0+\rho(\zeta_t\beta_0)+\mu_{it})) \tag{vii}$$

and can be approximated by

$$\Pr(B_j\nu_{it}\geq -B_j(X_{it}\widehat{\beta}_2+\rho(\widehat{\zeta}_t\widehat{\beta}_2)+\mu_{it})), \tag{viii}$$

where $\widehat{\zeta}_t=X_{it}-\widehat{X_{it}}$ and $\widehat{\beta}_2$ is the instrumental variables estimator of β_0. The simulated method of moments can be done now as above, where the parameters of the estimation are $V_\nu=\mathrm{Var}(\nu_{it})$ and ρ. We can then compute the variance of ε_{it} as $\rho\mathrm{Var}(\zeta_t\beta_0)\rho'+V_\nu$, where $\mathrm{Var}(\zeta_t\beta_0)$ can be consistently approximated by $\frac{1}{T}\sum_t\widehat{\zeta}_t\widehat{\beta}_2(\widehat{\zeta}_t\widehat{\beta}_2)'$. We can reduce the number of parameters to maximize over, if we use the following relationship between ρ and V_ν:

$$\begin{aligned} V_{\widetilde{\varepsilon}} &= \mathrm{Var}(\widetilde{\varepsilon}_{it})=\mathrm{Var}(\zeta_t\beta_0+\varepsilon_{it})=\mathrm{Var}(\zeta_t\beta_0+\rho\zeta_t\beta_0+\nu_{it}) \\ &= \mathrm{Var}((I+\rho)\zeta_t\beta_0+\nu_{it})=(I+\rho)\mathrm{Var}(\zeta_t\beta_0)(I+\rho)'+V_\nu. \end{aligned}$$

Hence,

$$V_\nu=V_{\widetilde{\varepsilon}}-(I+\rho)\mathrm{Var}(\zeta_t\beta_0)(I+\rho)' \tag{ix}$$

where $V_{\widetilde{\varepsilon}}$ was estimated during the instrumental variables estimation of β_0.

3. *Estimation Steps under the Null and with Instrumental Variables.*

Estimation under the null hypothesis is done through the following steps:

1. Fix θ_s – the initial value of the parameters. Calculate $I_{ijt} = \frac{\partial \log \widehat{P}_{ijt}(\theta_s)}{\partial \theta}$ for each (i, j, t) through Monte-Carlo simulation of the probability. Use independent draws for different (i, j, t).

2. Minimize $||g(\theta)||^2$ over θ, where $g(\theta) = \sum_{i,j,t}(d_{ijt} - \widehat{P}_{ijt}(\theta)) \cdot I_{ijt}$, in which $\widehat{P}_{ijt}(\theta)$ is obtained through Monte-Carlo simulation using draws independent of the draws in Step 1 above, and independent across (i, j, t), but without changing the draws across calculations of $g(\theta)$. Denote the point of minimum by $\widehat{\theta}$.

3. If $\widehat{\theta}$ is different by an amount larger than one expected standard deviation of any of the parameter estimates from θ_s used in Step 1 above, set θ_s to $\widehat{\theta}$ and go to Step 1. Changing θ_s twice in this fashion seems to be enough. Once θ_s is close to $\widehat{\theta}$, the later is our estimate of the parameters.

Estimation with instrumental variables is done through the following steps:

I. Change X_{ijt} to $\widehat{X}_{ijt}$ in the function for calculating choice probabilities P_{ijt} by taking the fitted values from OLS regression of X_{ijt} on instruments (see Eq. (21)).

II. Apply the estimation steps under the null (see Steps 1 to 3 above) with X_{ijt} changed to $\widehat{X}_{ijt}$ as calculated in Step I. The resulting $\widehat{\theta}$ is our estimate of the parameters with instrumental variables.

REFERENCES

AUTHOR (s), (2001) "Endogeneity and Individual Consumer Choice," *working paper,* university.

BASS, F. (1969), "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research,* **6**, 291-300.

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

BESANKO, D., S. GUPTA, AND D. JAIN (1998), "Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework," *Management Science,* **44**, 1533-1547.

BIJMOLT, T.H.A., H.J. VAN HEERDE, AND R. PETERS (2005), "New Empirical Generalizations on the Determinants of Price Elasticity," *Journal of Marketing Research,* **42**, 141-156.

BLATTBERG, R., G. EPPEN, AND J. LIEBERMAN (1981), "A Theoretical and Empirical Investigation of Price Deals for Consumer Nondurables," *Journal of Marketing,* **45**, 116-129.

BLUNDELL, R., AND J.L. POWELL (2004), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies,* **71**, 655-679.

BUCKLIN, R., S. GUPTA, AND S. HAN (1995), "A Brand's Eye View of Response Segmentation in Consumer Brand Choice Behavior," *Journal of Marketing Research,* **32**, 66-74.

CHAMBERLAIN, G. (1984), *Panel Data,* in Z. Griliches and M. Intriligator, eds., Handbook of Econometrics 2. Amsterdam: Elsevier Science Publishers.

CHINTAGUNTA, P.K., D. JAIN, AND N. VILCASSIM (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research,* **28**, 417-428.

CHINTAGUNTA, P.K., E. KYRIAZIDOU, AND J. PERKTOLD (2000), "Panel Data Analysis of Household Brand Choices," *working paper,* University of Chicago.

CHINTAGUNTA, P.K. (2001), "Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data," *Marketing Science,* **20**, 442-456.

CHINTAGUNTA, P.K., J.-P. DUBÉ, AND K.Y. GOH (2005), "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models," *Management Science,* **51**, 832-849.

GONUL, F., AND K. SRINIVASAN (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science,* **12**, 213-229.

GOOLSBEE, A., AND A. PETRIN (2004), "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica,* **72**, 351-381.

GROVER, R., AND V. SRINIVASAN (1992), "Evaluating the Multiple Effects of Retail Promotions on Brand Loyal and Brand Switching Segments," *Journal of Marketing Research,* **29**, 76-89.

GUADAGNI, P., AND J. LITTLE (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science,* **2**, 203-238.

GUPTA, S. (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research,* **25**, 342-355.

HAHN, J., AND J.A. HAUSMAN (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica,* **70**, 163-189.

HAJIVASSILIOU, V., D. MCFADDEN, AND P. RUUD (1996), "Simulation of Multivariate Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," *Journal of Econometrics,* **72**, 85-134.

HAUSMAN, J.A. (1978), "Specification Tests in Econometrics," *Econometrica,* **76**, 1251-1271.

HAUSMAN, J.A., AND P.A. RUUD (1987), "Specifying and Testing Econometric Models of Rank-Ordered Data," *Journal of Econometrics,* **34**, 83-104.

HONORÉ, B., AND E. KYRIAZIDOU (2000), "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica,* **68**, 839-874.

JEULAND, A., AND C. NARASIMHAN (1985), "Dealing – Temporal Price Cuts – by Seller as a Buyer Discrimination Mechanism," *Journal of Business,* **58**, 295-308.

KAMAKURA, W., AND G. RUSSELL (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research,* **26**, 379-390.

LUAN, J., AND K. SUDHIR (2006), "Forecasting Advertising Responsiveness for Short Lifecycle Products: The Case of DVDs," *working paper,* Yale University.

KIM, B.-D., K. SRINIVASAN, AND R. WILCOX (1999), "Identifying Price Sensitive Consumers: The Relative Merits of Demographic vs. Purchase Pattern Information," *Journal of Retailing,* **75**, 173-193.

LAL, R., AND C. MATUTES (1994), "Retail Pricing and Advertising Strategies," *Journal of Business,* **67**, 345-370.

MACÉ, S., AND S.A. NESLIN (2004), "The Determinants of Pre- and Postpromotion Dips in Sales of Frequently Purchased Goods," *Journal of Marketing Research,* **41**, 339-350.

MANCHANDA, P., P.E. ROSSI, AND P.K. CHINTAGUNTA (2004), "Response Modeling with Non-Random Marketing Mix Variables," *Journal of Marketing Research,* **41**, 467-478.

MCFADDEN, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica,* **57**, 995-1026.

McFadden, D., and K. Train (2000), "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics,* **15**, 447-470.

Moorthy, K.S. (1985), "Using Game Theory to Model Competition," *Journal of Marketing Research,* **22**, 262-282.

Morgan, M. (1990), *The History of Economic Ideas,* Cambridge: University Press, New York.

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

Newey, W. and D. McFadden (1994), *Large Sample Estimation and Hypothesis Testing,* in Z. Griliches and M. Intriligator, eds., Handbook of Econometrics 4. Amsterdam: Elsevier Science Publishers.

Petrin, A., and K. Train (2002), "Omitted Product Attributes in Discrete Choice Models," *working paper,* University of Chicago and University of California, Berkeley.

Rao, R. and N. Syam (2001), "Equilibrium Price Communication and Unadvertised Specials by Competing Supermarkets," *Marketing Science,* **20**, 61-81.

Rivers, D., and Q.H. Vuong (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics,* **39**, 347-366.

Rossi, P. and G. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research,* **30**, 171-182.

Roy, A., D.M. Hanssens, and J. Raju (1994), "Competitive Pricing by a Price Leader," *Management Science,* **40**, 809-823.

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

Sun, B., S.A. Neslin, and K. Srinivasan (2003), "Measuring the Impact of Promotions on Brand Switching When Consumers Are Forward Looking," *Journal of Marketing Research,* **40**, 389-405.

van Heerde, H., S. Gupta, and D. Wittink (2003), "Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is," *Journal of Marketing Research,* **40**, 481-491.

Villas-Boas, J.M., and R.S. Winer (1994), "Endogeneity in Brand Choice Models," *working paper,* University of California, Berkeley.

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

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

WOOLDRIDGE, J. (1994), *Estimation and Inference for Dependent Processes,* in R. Engle, D. McFadden, eds., Handbook of Econometrics 4. Amsterdam: Elsevier Science Publishers, pp. 2639-2738.

YANG, S., Y. CHEN, AND G.M. ALLENBY (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics,* **1**, 251-304.