Endogeneity in Brand Choice Models

J. Miguel Villas-Boas • Russell S. Winer
Haas School of Business, University of California at Berkeley, Berkeley, California 94720
villas@haas.berkeley.edu • winer@haas.berkeley.edu

Applications of random utility models to scanner data have been widely presented in marketing for the last 20 years. One particular problem with these applications is that they have ignored possible correlations between the independent variables in the deterministic component of utility (price, promotion, etc.) and the stochastic component or error term. In fact, marketing-mix variables, such as price, not only affect brand purchasing probabilities but are themselves endogenously set by marketing managers. This work tests whether these endogeneity problems are important enough to warrant consideration when estimating random utility models with scanner panel data. Our results show that not accounting for endogeneity may result in a substantial bias in the parameter estimates.
(*Endogeneity; Brand Choice Models; Panel Data; Common Demand Shocks; Limited Dependent Variables*)

1. Introduction

Over the past 20 years, there has been a significant increase in the research literature on the topic of modeling the choice process of economic agents or consumers. Motivated by McFadden's (1973) work on the conditional multinomial logit model, there has been particular interest in what are referred to as random utility models where the decision maker faces a choice set for which the utility of each alternative is a random variable. This utility is usually specified to have a deterministic component which is a function of observable variables and a random component for which a variety of parametric assumptions have been made. In particular, random utility models, mainly logit and probit models, have been often applied to household purchasing behavior using electronic scanner data (e.g., Guadagni and Little 1983).

One potential problem in how marketing researchers have applied these random utility models to scanner data is that they have ignored the fact that the exogenous deterministic components of utility such as price, promotion, advertising, etc., are themselves endogenous. Marketing managers set these marketing-mix variables based on market information which may be in part unobservable to the researcher but which nevertheless affects consumer choice. This would create a situation where the marketing-mix variables could be correlated with the error terms in the latent utilities.[1]

The term "endogeneity" has been used in the literature (see some of the papers referred to in this paper) and results from both the marketing-mix variables and the purchase decisions being determined simultaneously as a function of the market situation which is unobservable to the researcher. This "endogeneity" can be stated as an "errors in variables" problem in the econometric model.

Failure to account for this endogeneity in the deterministic components of utility has the potential to bias the parameter estimates of the marketing-mix variables. If this happens, both major uses of these models, for diagnostic and optimization purposes, could produce misleading results, seriously affecting the out-

[1] A second source of errors in variables problems in scanner data applications, not considered here, is the way the price variable has been constructed to include coupon redemption. Coupon redemption is also likely to be strongly correlated with the error term (in particular, with the heterogeneity disturbance in any formulation accounting for consumer heterogeneity).

Management Science/Vol. 45, No. 10, October 1999
pp. 1324–1338

0025-1909/99/4510/1324$05.00

comes of the marketing decisions.[2] This problem of endogeneity is not, of course, a new topic in the marketing literature as Bass (1969) discussed the problem in the context of the simultaneity issue between advertising and sales.

The justification researchers have presented for not accounting for these endogeneity problems in brand choice models with scanner data has to do with the fact that, in a certain period, the marketing-mix variables are common across all consumers. If the utilities' error terms are independent across individuals, it is unlikely that the marketing-mix variables (common across all individuals) are highly correlated with the error terms. However, in this paper, we question this independence assumption and show that it may not hold. This may result in a correlation between the error terms and the marketing-mix variables.

Other researchers have pointed out that another justification for not considering the endogeneity issue when estimating the parameters of an individual choice model is that individuals are price takers and therefore, strategically do not impact the price setting behavior of the sellers. However, individuals being price takers is not an argument for endogeneity not to be a serious problem when estimating an individual choice model: price-taking individuals still make decisions as a function of variables that are unobservable to the researcher, that may be observable by the price-setting firms, and which may affect the firms' pricing behavior. If this (likely) event occurs, endogeneity will still be a potential problem.

Why may the error terms in the latent utilities not be independent across individuals? One possibility is that there are idiosyncratic word-of-mouth or fashion (wearout) effects (possibly positively correlated through time). Alternatively, there are always market phenomena that affect all households and are not observable by the researcher (and, therefore, not included in the choice model), but that the marketing managers use in their decisions. These are referred to in what follows as common demand "shocks." These effects are exactly the same ones that are present in endogeneity studies on aggregate demand. For example, the common demand shocks can represent difficult to quantify aspects of style, prestige, reputation, past experience, or quantifiable aspects about which the researcher does not have information (Berry et al. 1995). In addition, the common shocks can represent taste changes induced by other marketing-mix variables about which the researcher does not have information such as in-store effects, advertising, or coupon availability (Besanko et al. 1998). Another example of such a phenomenon is significant activity in a competing product class that benefits or hurts some of the brands in the category under study. A health concern about beef may affect the demand for the leading brand of ketchup, Heinz.

The major purpose of this work is to test whether these endogeneity problems are sufficiently important to warrant consideration when estimating random utility models with scanner panel data. In particular, we focus on testing endogeneity of the pricing decision. While testing for endogeneity, we also obtain consistent estimates of the parameters of the random utility model under the assumption that this model is well specified. Several theoretical tests for endogeneity in models of limited dependent variables have been presented by, among others, Heckman (1978), Newey (1985, 1987), and Rivers and Vuong (1988). Implementations of these tests have not been very frequent (one example is Heckman 1979) and we are unaware of any applications in the context of brand choice models (where an explanatory variable, such as price, has the same value for several individuals).

The tests are performed in two frequently purchased product categories: yogurt and ketchup. We find that, subject to the assumptions of the utility model, the bias introduced by ignoring endogeneity can be substantial in that there are important changes in the measures of price sensitivity and the individual brand alternative parameters. We also show that the base model results are robust to different specifications of the error structure, to different specifications of the deterministic component of the random utility model, and to different instruments for the decision variables.

The paper is organized as follows. In §2 we

[2] It is still possible that these models can be used for prediction purposes if the managers keep making decisions in the same way as in the data used for estimation.

present the results of a base model that summarizes the main message of the paper. This model has a very simple error structure, considers endogeneity on the additive term of the random utility model, and includes state dependence as a lagged purchase indicator. In this section we also report results on the use of different instruments for price, and on the use of a loyalty variable as in Guadagni and Little (1983). For the more interested or skeptical reader, §3 presents several robustness checks and the extension to more general error structures. Section 4 concludes and discusses directions for future research.

2. A Base Model with Endogeneity

This section considers a base random utility model with endogeneity. This model captures the main effects that we want to present in this paper. The model is kept as simple as possible in order to clearly illustrate the endogeneity effects. Later in this section and in the following section we present and report results of several extensions of the base model.

Consumers are assumed to buy one product from a choice set C with J alternatives. A consumer derives utility only from buying one product in that choice set.[3] The consumer buys the product for which the perceived utility minus the price (which we call the indirect utility) is the greatest. In this section, utility has a deterministic component and a random component. The deterministic component has a fixed term, a signaling effect term (with the variables feature and display; see, for example, Milgrom and Roberts 1986), and a lagged purchase indicator as a way of accounting for heterogeneity among consumers. We have also performed runs including either a loyalty variable specified as an exponential smoothing of past purchases (as in Guadagni and Little 1983), or unobserved heterogeneity, and the main results remain unchanged (Villas-Boas and Winer 1994). In this base model, the lagged purchase indicator equals one if the consumer purchased that product the last time she purchased any product in this category, and is set to zero otherwise.

Given that all that matters in the decision of which product to choose is which one yields the highest indirect utility, the form described below is unique up to a monotonic transformation. We can write the indirect utility function as

$$U_{ijt} = X'_{ijt}\beta + \epsilon^a_{ijt} + \epsilon^b_{jt} \tag{1}$$

where i indexes households, j represents the brands, and t represents the week in which household i purchased a product from the choice set C (the set of consumers that made a purchase in week t is referred to as I_t; the product chosen by consumer i in week t is referred to as j_{it}). X_{ijt} is a vector with the following variables: dummy variables for the number of alternatives minus one, price, display, feature, and lagged purchase indicator.[4] β is the vector of parameters to be estimated. U_{ijt} is a latent variable; the researcher observes the product being chosen, which we call d_{it}. The error term ϵ^a_{ijt} is assumed to be Gumbel distributed with parameters $(0, \theta)$.[5] The error term ϵ^b_{jt} is assumed to be normally distributed with mean zero and variance $\sigma^2_{\epsilon^b}$. ϵ^b_{jt} represents market phenomena or brand-specific variables that affect all households and are not observable by the researcher (and, therefore, not included in the choice model).

Furthermore, it is assumed that all the elements of the set $\{\epsilon^a_{ijt}, \epsilon^b_{jt}\ \forall i, j, t\}$ are independent from each other (in the next section we allow for more general error structures). The scale of the indirect utility is set by $\sigma^2_{\epsilon^b} + \pi^2/(6\theta^2) = (\pi^2/6)$.

The main difference from the usual models is the existence of ϵ^b_{jt} which represents the common demand

[3] In those weeks the consumer has utility equal to minus infinity if she does not buy any product, and the utility does not improve if she buys more than one product. We have also performed runs of the model with the no-purchase option: The results are similar to the ones presented here, and our main message is substantially strengthened.

[4] More information about the individual consumers was not available. If available, this information should be used in the analysis. One may argue that it is unlikely to affect the results below because it is, by definition, part of the individual specific residual and not the common residual.

[5] This is for easier computability. In the following section we consider ϵ^a_{ijt} to be normally distributed.

shocks of the type described in the Introduction.[6] If $\sigma^2_{\epsilon^b} = 0$, this model reduces to the usual models. Testing for common demand shocks is testing for the null hypothesis that $\sigma^2_{\epsilon^b} = 0$.

We allow for endogeneity in prices. In order to do this, we use instruments, W, for prices such that

$$P_{jt} = p_j(W_{jt}; \alpha) + \eta_{jt} \tag{2}$$

where P_{jt} is the price set for brand j in period t, and $p_j(\ )$ is a known functional form of W_{jt} with parameters α. Using instruments for the endogenous variables, without taking into account any restrictions of the parameters across equations, is known as the "limited information" approach. The main advantage of this approach is that specification problems in the pricing equation do not extend to the parameters of the utility equation—the main objective of the analysis.[7]

In order to obtain consistent estimates of the parameters of the utility equation we then simply need the instruments W to be correlated with the prices and independent of the errors in the utility equation, in particular, independent of ϵ^b_{jt}. Any instruments (variables) that satisfy these two conditions generate consistent estimates of the parameters of the utility equation, subject to the utility equation being correctly specified (see Rivers and Vuong 1988).[8] These variables are called *instruments* because we are using its assumed relationship to the price variables and the common errors in the brand choice equation to be able to identify the price coefficient in the brand choice equation (e.g., Green 1997, p. 288).[9]

Note also that this estimation method is sufficiently general such that the current price may result from both retailers and manufacturers activities (or just one or several retailers, or just manufacturers). In the paper we use the term "seller" or "firm" to represent the decision maker for one of the prices, but the formulation is sufficiently general such that it could be the same entity deciding on all the prices in the set C.[10]

An easily available set of instruments for the prices are simply lagged prices. We use these instruments in most of what follows. The pricing equation then becomes

$$P_{jt} = \alpha_{j0} + \alpha_{j1}P_{j,t-1} + \eta_{jt}. \tag{3}$$

Note that because of forward buying and stock piling (or other reasons), one might argue that lagged prices could be correlated with the utility equation error terms, which would make lagged prices not appropriate instruments. However, these forward buying and stock piling effects have been found to be of very limited size (see Gupta 1988). Prices could still be correlated through time because of cost correlation through time. Another potential and more serious problem is that the common shock in the utility equation, ϵ^b_{jt}, may be correlated through time, and then lagged prices would be correlated with the current period common shock. But, if this is the case, the results obtained here would be conservative and would only underestimate the endogeneity effects. A main advantage of using lagged prices as instruments is that they are readily available to the researcher. Below we report results from the use of other instruments for price (immune to any potential correlation between lagged prices and the latent utilities error terms) that confirm and strengthen the results obtained with lagged price as the only instrument.

Note that the point of this pricing equation is not to model price-setting behavior but it is only a robust way of uncovering the true parameters of the random utility model (instrumental variables estimation—see Rivers and Vuong 1988). However, one can envision a market situation where this pricing equation could be derived from the equilibrium profit maximization (of the retailer, manufacturers, or both) if the costs are

[6] The formulation for the common demand shocks is a random effects one because of the small number of purchases per week. An alternative is a fixed effects formulation.

[7] See Tellis and Zufryden (1995) for a derivation of optimal pricing rules that could be used in a "full information" approach. See also Besanko et al. (1998) and Kadiyali et al. (1995) for examples of the "full information" approach with aggregate data.

[8] A proof of this statement in the context of this paper is available upon request from the authors.

[9] The methods using instruments are very general, and include several variations of two-stage least squares to methods involving maximum likelihood, and, in particular, limited information maximum likelihood (LIML) which is used in this paper.

[10] In fact, if we restrict attention to the brands in one retailer, the prices are decided by a combination of that retailer and the manufacturers' decisions.

correlated through time and the firms observe only a certain part of the demand shocks.[11] Alternatively, the pricing equation could be simply seen as a rule justified by bounded rationality on the part of the firms' decision makers.

The error η_{jt} consists of unobservables (to the researcher) which reflect shocks on costs (for example, productivity, input prices) and on demand (part of ϵ^b_{jt}). Because η_{jt} consists in part of demand shocks, it is very likely that η_{jt} is correlated with ϵ^b_{jt}. Firms set different prices through time, not only because their costs vary, but also because demand characteristics (unobserved to the researcher but partially observed by the firms) vary. This is the dimension of the endogeneity problem that is studied in this paper.

The argument can be illustrated with the following stylized example. Consider a monopolistic market with demand function $D(P_t) = a - bP_t + \epsilon_t$ where ϵ_t is not observed by the researcher. Suppose that the monopolist observes ϵ_t perfectly and has marginal cost c_t. Then, if the monopolist maximizes profits, it will set $P_t = [(a + bc_t + \epsilon_t)/(2b)]$, and P_t and ϵ_t will be positively correlated because $(dP_t/d\epsilon_t) = (1/(2b))$. This correlation is small in absolute value if the variance of c_t is much greater than the variance of ϵ_t. In the real world (i) most of markets we are interested in are not monopolies, (ii) firms may not be fully rational or maximize profits, or (iii) the firms are not able to perfectly observe the demand shock. However, we should still expect to observe some correlation between prices and the residuals in the demand equation. With individual data, the same problem arises if we have a common residual (shock) across individuals.

The correlation between the household level error term and the aggregate price model error term might be small if there is a large variance in costs, if the signal received by the firms of the households' common shock is not very accurate, or if the variability of the household specific shocks is much greater than the variability of the households' common shock. This is an empirical question to which we provide an answer.

[11] Solving explicitly for the retail prices equilibrium one obtains equations with the retail prices on the right-hand side and error terms plus exogenous variables on the left-hand side.

There is a long literature in econometrics addressing the issue of possible correlation between the household error term and the price model error term. For an history of this literature in econometrics see Morgan (1990, Part 2, Chapter 6). More recently, closer to the framework in this paper, with a full information approach, and in independent or posterior work, see Berry (1994), Berry et al. (1995), and Besanko et al. (1998). All these papers use aggregate data while our paper uses household data. If household data is available, it should be used because it has more information and may allow for a more accurate testing of theories of consumer behavior.

We assume that η_{jt} is normally distributed, with mean zero, and variance σ^2_η, and that $E[\eta_{jt}\epsilon^b_{jt}] = \rho\sigma_{\epsilon^b}\sigma_\eta$.[12] Testing for the endogeneity problem to be significant is testing for the null hypothesis that $\rho = 0$.

It is also assumed in this section that

$$E[\eta_{jt}\eta_{j't'}] = 0 \quad \text{if } j \neq j' \text{ or } t \neq t', \tag{4}$$

$$E[\eta_{jt}\epsilon^b_{j't'}] = 0 \quad \text{if } j \neq j' \text{ or } t \neq t', \tag{5}$$

and that all the elements of the set $\{\epsilon^a_{ijt}, \eta_{jt}\ \forall i, j, t\}$ are independent from each other.

Assuming η_{jt} to be independent across j requires, given existing game-theoretic models of interaction among firms, that the firms are only able to observe each of their own demand shocks and not the competitors' demand shocks. This is relaxed below.[13]

[12] Because prices are always positive, the normality assumption on η_{jt} can be misspecified, and create bias in the estimation. However, in our applications the estimates of σ^2_η are much smaller (about 20 times smaller) than the average price for the constraint "positive prices" to have any significant impact on the bias of the estimates. We also used estimation methods that were robust to the marginal distribution of η_{jt} and the main results were unchanged.

[13] For the profit maximization interpretation of the pricing equation, the assumption that η_{jt} is independent across t may also not be empirically satisfied. This assumption is kept throughout the paper. Relaxing it would create serious estimation problems in the sense that the likelihood of the sample would be a nonseparable probability, i.e., the log-likelihood function would not be additively separable across periods. Note, however, that if the the process of $(\eta_{jt}, \epsilon^b_{jt}, \epsilon^a_{ijt})$ in t is stationary and ergodic, the parameter estimators described here are still consistent, and the method we use would be a quasi-maximum likelihood method (see Wooldridge 1994).

This model can be estimated using maximum likelihood. The likelihood function can be written as

$$L(\beta, \alpha_0, \alpha_1, \sigma^2_{\epsilon^b}, \sigma^2_\eta, \rho) = \prod_{t=1}^{T} \prod_{h=1}^{J} f(\eta_{ht}; \sigma^2_\eta; \alpha_0, \alpha_1) g(d_{it} \ \forall\, i \in I_t | \eta_{jt} \ \forall\, j) \quad (6)$$

where α_0 and α_1 are vectors with generic elements, respectively, α_{j0} and α_{j1}, $f(\eta_{ht}; \sigma^2_\eta; \alpha_0, \alpha_1)$ is the density function of a normal random variable with mean zero and variance σ^2_η, and $g(d_{it} \ \forall i \in I_t | \eta_{jt} \ \forall j)$ is the probability of observing choices d_{it} $\forall i \in I_t$ in week t given η_{jt} $\forall j \in J$. Finally, $g(d_{it} \ \forall i \in I_t | \eta_{jt} \ \forall j)$, the conditional probability of observing the purchases in week t given the shocks in the pricing equation, can be written as

$$g(d_{it} \ \forall\, i \in I_t | \eta_{jt} \ \forall\, j) = \int \int \cdots \int \prod_{i \in I_t} \frac{e^{\theta(X^i_{j_{it}t}\beta + \epsilon^b_{j_{it}t})}}{\sum_{j=1}^{J} e^{\theta(X^i_{jt}\beta + \epsilon^b_{jt})}} \times \prod_{k=1}^{J} f(\epsilon^b_{kt} | \eta_{kt}; \rho, \sigma^2_{\epsilon^b}, \sigma^2_\eta) d\epsilon^b_{kt} \quad (7)$$

where $f(\epsilon^b_{kt} | \eta_{kt}; \rho, \sigma^2_{\epsilon^b}, \sigma^2_\eta)$ is the conditional density distribution of the normally distributed random variable ϵ^b_{kt} on the normally distributed random variable η_{kt}, both with zero mean, with correlation ρ, and variances, respectively, $\sigma^2_{\epsilon^b}$ and σ^2_η.[14]

If endogeneity is important (i.e., $\rho \neq 0$) one might expect prices to be positively correlated with the error terms of the latent utilities (i.e., with greater demand firms set higher prices) which results in $\rho > 0$. Furthermore, because of this positive correlation, we expect that the failure to account for endogeneity in the choice model results in the price effect being underestimated.

This model was estimated using scanner panel data from both the yogurt and the ketchup markets. In the yogurt market, we restricted our attention to C = {Dannon, Yoplait, Private Label} (Dannon and the Private Label were in 8 oz. packages; Yoplait was in a 6 oz. package); these brands account for 77% of the market. There are 3,513 purchase occasions on C in the data being used. The purchases occur during 137 weeks. We restrict our attention to only the largest store in order to simplify the analysis.[15] In the ketchup market we restricted our attention to C = {DelMonte, Hunts, Heinz} for the 32 oz. package, which accounts for 57% of the market.[16] There are 1,256 purchase occasions on C in the data being used. The purchase occasions occur during 129 weeks. We restrict our attention to only one store of the three largest.[17] All data on prices, displays, and features are recovered from the purchase occasions files. The results are presented in Table 1.

These results clearly show the importance of considering the endogeneity of the pricing decision. First, there are important common shocks across individuals as the hypothesis $\sigma_{\epsilon^b} = 0$ is rejected for both product categories. Note that the value of the chi-square statistic on this restriction is 51.6 for the yogurt market and 33.0 for the ketchup market (the critical value at the 5% significance level is 3.84).[18] Second, there is significant endogeneity of the price variable as the hypothesis $\rho = 0$ is rejected for both product categories. Note that the chi-square statistic is 36.8 for the yogurt market and 5.0 for the ketchup market (the critical value at the 5% significance level is also 3.84).

[14] Another estimation method would be to substitute the pricing equation into the latent utility equation. However, this method is inefficient because the information on current prices is not used and, furthermore, does not allow the researcher to compute the demand sensitivity to price, one of the main necessary elements for the diagnostic and optimization purposes of these models.

[15] There were originally 10 stores in this data set. Including more stores in the analysis would only increase the choice set into products where there was a small degree of substitutability for the consumers being considered. The analysis reported in this paper was limited to one store for both the yogurt and the ketchup markets. The results for more than one store are in every way similar to the results presented here and available upon request from the authors.

[16] We did not include more brands in the analysis for computational reasons.

[17] There were 9 stores in this data set.

[18] The value of the maximum log-likelihood can be positive in these cases because the density of η_{jt}, included in the log-likelihood function, can be greater than one (i.e., the log of the density can be greater than zero), and compensate for the log of the probability of choice of each brand (which is less than zero).

Table 1 Results from the Base Case

	Complete Model	$\rho = 0$	$\rho = 0, \sigma_{\epsilon^b} = 0$
Yogurt Market			
Dannon	5.67	3.48	3.32
	(0.07)	(0.08)	(0.12)
Yoplait	5.75	3.52	3.34
	(0.10)	(0.08)	(0.13)
Price	−21.74	−14.22	−13.83
	(0.20)	(0.31)	(0.45)
Display	0.33	0.51	0.54
	(0.03)	(0.04)	(0.09)
Feature	−0.015	−0.147	−0.069
	(0.02)	(0.02)	(0.07)
LPI	1.63	1.75	1.79
	(0.03)	(0.04)	(0.04)
σ_η	0.039	0.038	0.038
	(0.001)	(0.002)	(0.002)
σ_{ϵ^b}	0.61	0.40	
	(0.04)	(0.09)	
ρ	0.78		
	(0.05)		
LL	−1086.2	−1104.6	−1130.4
Ketchup Market			
DelMonte	−2.89	−2.54	−2.46
	(0.13)	(0.16)	(0.26)
Hunts	−2.69	−2.43	−2.40
	(0.13)	(0.15)	(0.25)
Price	−6.21	−5.61	−5.58
	(0.21)	(0.31)	(0.56)
Display	1.22	1.16	1.14
	(0.11)	(0.10)	(0.07)
Feature	0.80	0.76	0.83
	(0.12)	(0.10)	(0.10)
LPI	1.24	1.25	1.25
	(0.07)	(0.07)	(0.08)
σ_η	0.056	0.056	0.056
	(0.001)	(0.001)	(0.001)
σ_{ϵ^b}	0.56	0.50	
	(0.08)	(0.09)	
ρ	0.30		
	(0.09)		
LL	97.6	95.1	78.6

Note: LL is the value of the log likelihood without the fixed terms; the standard errors are in parentheses; LPI stands for lagged purchase indicator (this is also the case for all tables).

Third, the parameter that estimates the impact of price on choice seems to be underestimated (expected direction of the bias) if endogeneity is not taken into account. For the yogurt market the price effect is underestimated by 57%; for the ketchup market, it is underestimated by 11%. This point difference is not statistically significant for the ketchup market, but, as argued below, this seems to be due to the lagged price instruments.

Note also that the correlation between ϵ_{jt}^b and η_{jt}, ρ, is positive as expected (it is quite large for the yogurt market). The common demand shocks are 23% of the total demand shocks for the yogurt market, and 19% for the ketchup market. Finally, note that there is a pattern in the change of the estimates of the parameters of the latent utility model when the endogeneity of the pricing decision is allowed for: the brand-specific dummies increase in absolute value, and the effect of the other marketing-mix variables and of the lagged purchase indicator decrease (for most of the coefficients).

These results present the main message of the paper: accounting for the endogeneity of a marketing-mix variable in a brand choice model using household data can result in a significant change in the parameter estimates. This is important because changes in parameter estimates may then affect the resulting optimal resource allocation. As argued above, this result is also important because it shows that using household data, and for some markets, individual uncertainty is not much more important than the common uncertainty, and therefore, endogeneity can still be an important problem.

It is well known that in this type of models the elasticities of demand with respect to the explanatory variables are roughly (exactly for the standard logit model) proportional to the coefficients of the deterministic component of the latent utility. To illustrate this situation in this setting, Table 2 reports the point estimates of the own-price elasticities at the sample mean. The elasticities for the traditional model are somewhat higher than other similar findings (Guadagni and Little 1983, Tellis 1988) for several reasons. First, our price variable includes promotions and thus reflects the impact of both marketing mix variables. Second, since our data are for only one store, the unique demand conditions of that single retail outlet significantly affects the estimates. Finally, our esti-

Table 2 Own-Price Elasticities for Base Case

	Complete Model	Traditional Model
Yogurt Market		
Dannon	−10.50	−6.94
Yoplait	−11.91	−7.55
Private Label	−3.37	−1.98
Ketchup Market		
DelMonte	−2.43	−2.12
Hunts	−4.50	−4.12
Heinz	−7.14	−6.37

Note: Elasticities are computed at the sample mean.

mates use household data as opposed to Tellis's meta-analysis on largely aggregate studies. Clearly, however, the differences in elasticities between our model and the traditional restricted model indicate that optimal decisions would be significantly impacted by omitting the endogeneity problem, particularly for brands in the yogurt category.

In order to alleviate the computational problems when using maximum likelihood estimation, we use in the remainder of the paper the method of simulated maximum likelihood (SML).[19] For some simulators, SML is equivalent to the method of simulated scores (see Keane 1993 and Hajivassiliou and McFadden 1993, and Gourieroux and Monfort 1993 for a discussion of the properties of these estimators).[20] Because it is an expected value of a smooth function, we simulated $g(\cdot)$ by taking the mean across draws of ϵ_{jt}^{b}. In order to check the accuracy of SML we first simulated $g(\cdot)$ for the model above. The results (not presented here but available upon request from the authors) are very similar to the results presented in the Table 1.

In order to check the robustness of the results to other instruments for price, we also allowed Equation (2) to include other explanatory variables. One pricing equation that has been used in the past in several applications (for example, Winer 1986), is to have both lagged market share and lagged price variables in the pricing equation.

Equation (2) is then changed to

$$P_{jt} = \alpha_{j0} + \alpha_{j1}P_{j,t-1} + \alpha_{j2}S_{j,t-1} + \eta_{jt} \qquad (8)$$

where S_{jt} is the market share for brand j in period t, and α_{j2} is a parameter. The rationale behind this model is that managers adjust their prices based on their market performance in the prior period.

The results (see Villas-Boas and Winer 1994) show that the main conclusions are robust to including lagged market share as an explanatory variable: the common shocks across consumers are statistically significant and there is endogeneity in the price variable.

Another important alternative to consider is to eliminate lagged prices from the set of instruments and use only other variables as instruments for current prices. We had available some cost variables for both markets that could serve as instruments. These variables are more likely to be independent of the common demand shocks, and in this sense, be better instruments for price.[21] For yogurt, the cost variables were prices of milk (used to make yogurt) in different regions of the country. For ketchup the cost variables were prices of different types of tomatoes.[22] The cost variables had statistically significant explanatory power on current prices. The results, available upon request from the authors, strengthened, as expected, the main message of the results presented above: not accounting for endogeneity may result in a substantial underestimation of the price coefficient. The strengthening of the results

[19] This method allows also for estimation with any number of brands. Notice also that the standard method of simulated moments does not work in this case because the number of alternatives in that method is too large—J raised to the power of the number of consumers that purchased a product in that week (see Hajivassiliou and McFadden 1993).

[20] For the GHK simulator, mentioned below, these estimators are consistent and asymptotically normal if $K/\sqrt{N} \to \infty$ as $N \to \infty$, where K is the number of simulations and N is the number of observations.

[21] An alternative set of instruments would be to use the features of substitute products as in Bresnahan et al. (1997), but this option brings also additional issues of the same type as lagged prices.

[22] All cost variables were obtained from the United States Department of Agriculture.

is expected because, as discussed above, lagged prices are possibly still correlated with the current utility equation error terms. Furthermore, the ketchup results became much closer to the yogurt results in terms of endogeneity bias, which seems to indicate that the differences between the two cases in Table 1 is due to lagged prices being a less good instrument in the ketchup than in the yogurt market, and not due to a different degree of endogeneity bias in the two markets, i.e., not due to a different percentual bias in the price coefficient when not accounting for endogeneity. This is also consistent to the common belief expressed by some marketing executives contacted by us that prices vary less from period to period in the ketchup market than in the yogurt market. This smaller variation of prices through time in the ketchup market could be due to a strong positive correlation on the common errors, which would mean that lagged prices in the ketchup market would not be a completely satisfactory instrument (because they would be correlated with the current period common shocks).

3. More General Error Structures

In this section we consider the case of more general error structures. In particular, we consider structures where the independence of irrelevant alternatives property does not hold and where there might be nonzero correlations across j between the elements of the set $\{\epsilon_{jt}^b, \eta_{jt}\ \forall j, t\}$.

In order to allow the independence of irrelevant alternatives property not to hold, we assume ϵ_{ijt}^a to be normally distributed instead of Gumbel (as stated above we are now using a SML estimation method).

Then

$$g(d_{it}\ \ \forall\, i \in I_t | \eta_{jt}\ \ \forall\, j) = \int\!\!\int \cdots \int \prod_{i \in I_t} \text{Prob}(j_{it};\, i,\, t,\, \epsilon_{jt}^b\ \ \forall\, j,\, t) \cdot \prod_{k=1}^{J} f(\epsilon_{kt}^b | \eta_{kt};\, \rho,\, \sigma_{\epsilon^b}^2,\, \sigma_\eta^2)\, d\epsilon_{kt}^b \tag{9}$$

where

$$\begin{aligned}
&\text{Prob}(j_{it};\, i,\, t,\, \epsilon_{jt}^b\ \ \forall\, j,\, t) \\
&\quad = \tilde{\Phi}[(X'_{ij_{it}t} - X'_{i1t})\beta \\
&\qquad + \epsilon_{j_{it}t}^b - \epsilon_{1t}^b, \ldots, (X'_{ij_{it}t} - X'_{i(j_{it}-1)t})\beta \\
&\qquad + \epsilon_{j_{it}t}^b - \epsilon_{(j_{it}-1)t}^b, (X'_{ij_{it}t} - X'_{i(j_{it}+1)t})\beta \\
&\qquad + \epsilon_{j_{it}t}^b - \epsilon_{(j_{it}+1)t}^b, \ldots, (X'_{ij_{it}t} - X'_{iJt})\beta \\
&\qquad + \epsilon_{j_{it}t}^b - \epsilon_{Jt}^b]
\end{aligned} \tag{10}$$

where $\tilde{\Phi}[\,\cdot\,]$ is the cumulative normal distribution of dimension $J - 1$ of a vector with mean zero and variance-covariance the variance-covariance of the vector $(\epsilon_{i1t}^a - \epsilon_{ij_{it}t}^a, \ldots, \epsilon_{i(j_{it}-1)t}^a - \epsilon_{ij_{it}t}^a, \epsilon_{i(j_{it}+1)t}^a - \epsilon_{ij_{it}t}^a, \ldots, \epsilon_{iJt}^a - \epsilon_{ij_{it}t}^a)$. This can be obtained, for example, from the variance-covariance of the vector $(\epsilon_{i2t}^a - \epsilon_{i1t}^a, \epsilon_{i3t}^a - \epsilon_{i1t}^a, \ldots, \epsilon_{iJt}^a - \epsilon_{i1t}^a)$. The elements of this matrix are parameters of the likelihood function. In this particular case, because $J = 3$, the parameters are

$$V = \text{VAR}\begin{bmatrix} \epsilon_{i2t}^a - \epsilon_{i1t}^a \\ \epsilon_{i3t}^a - \epsilon_{i1t}^a \end{bmatrix} = \begin{bmatrix} \sigma_{\epsilon_{2-1}^a}^2 & \rho_{\epsilon_{2-1}^a \epsilon_{3-1}^a} \sigma_{\epsilon_{2-1}^a} \sigma_{\epsilon_{3-1}^a} \\ & \sigma_{\epsilon_{3-1}^a}^2 \end{bmatrix} \tag{11}$$

where now the indirect utility is scaled by $\sigma_{\epsilon_{2-1}^a}^2 + 2\sigma_{\epsilon^b}^2 = 1$.[23] It is clear from this specification that ϵ_{ijt}^a are not independent across j, i.e., the independence of irrelevant alternatives property may not hold.

Using SML in this model requires simulation of both $\tilde{\Phi}[\,\cdot\,]$ (draws of ϵ_{ijt}^a) and $g(\cdot)$ given $\tilde{\Phi}[\,\cdot\,]$ (draws of ϵ_{jt}^b). The simulation of $\tilde{\Phi}[\,\cdot\,]$ was performed using the GHK (Geweke-Hajivassiliou-Keane) simulator (see, for example, Hajivassiliou et al. 1992 for a description of the simulator). The simulation of $g(\cdot)$ was described above.

The results (still assuming η_{jt}, and ϵ_{jt}^b to be identically distributed and independent across j) are presented in Table 3 for the yogurt and the ketchup

[23] Note that this is the standard scaling in Probit models, which is different from the scaling used in the previous section.

Table 3 Simulation of $g(\cdot)$ and $\tilde{\Phi}[\,\cdot\,]$

size;7q

	Complete Model	$\rho = 0$	$\rho = 0,\ \sigma_{\epsilon^b} = 0$	$\rho_{\epsilon^a_{2-1}\epsilon^a_{3-1}} = 0$ $\sigma_{\epsilon^a_{2-1}} = \sigma_{\epsilon^a_{3-1}}$	$\rho_{\epsilon^a_{2-1}\epsilon^a_{3-1}} = 0$ $\sigma_{\epsilon^a_{2-1}} = \sigma_{\epsilon^a_{3-1}}$ $\rho = 0$	$\rho_{\epsilon^a_{2-1}\epsilon^a_{3-1}} = 0$ $\sigma_{\epsilon^a_{2-1}} = \sigma_{\epsilon^a_{3-1}}$ $\rho = 0,\ \sigma_{\epsilon^b} = 0$
Yogurt Market						
Dannon	2.69	1.60	1.49	3.90	2.05	1.87
	(0.28)	(0.11)	(0.08)	(0.23)	(0.10)	(0.06)
Yoplait	2.79	1.70	1.56	3.73	1.78	1.60
	(0.29)	(0.11)	(0.09)	(0.26)	(0.11)	(0.07)
Price	−10.55	−6.81	−6.41	−14.24	−7.84	−7.27
	(0.96)	(0.38)	(0.30)	(0.78)	(0.36)	(0.21)
Display	0.20	0.23	0.24	0.23	0.31	0.33
	(0.07)	(0.08)	(0.05)	(0.07)	(0.10)	(0.06)
Feature	0.027	−0.033	−0.016	0.025	−0.065	−0.039
	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.04)
LPI	0.80	0.83	0.85	0.83	1.00	1.01
	(0.04)	(0.04)	(0.04)	(0.05)	(0.02)	(0.02)
σ_η	0.039	0.038	0.038	0.040	0.038	0.038
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
σ_{ϵ^b}	0.29	0.16		0.42	0.20	
	(0.04)	(0.02)		(0.04)	(0.02)	
ρ	0.84			0.94		
	(0.05)			(0.02)		
$\rho_{\epsilon^a_{2-1}\epsilon^a_{3-1}}$	0.68	0.71	0.67			
	(0.03)	(0.03)	(0.03)			
$\sigma_{\epsilon^a_{3-1}}$	0.83	0.90	0.90	0.80	0.96	1.00
	(0.04)	(0.04)	(0.04)			
LL	−1093.8	−1110.4	−1133.6	−1167.3	−1198.8	−1216.1
Ketchup Market						
DelMonte	−1.29	−1.18	−1.10	−1.31	−1.17	−1.03
	(0.13)	(0.12)	(0.11)	(0.17)	(0.16)	(0.12)
Hunts	−1.29	−1.20	−1.16	−1.26	−1.15	−1.07
	(0.12)	(0.12)	(0.10)	(0.16)	(0.15)	(0.12)
Price	−2.48	−2.25	−2.19	−3.24	−2.96	−2.79
	(0.28)	(0.27)	(0.24)	(0.34)	(0.32)	(0.26)
Display	0.52	0.53	0.50	0.69	0.70	0.66
	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.04)
Feature	0.41	0.36	0.42	0.41	0.37	0.45
	(0.08)	(0.07)	(0.05)	(0.10)	(0.09)	(0.05)
LPI	0.59	0.59	0.60	0.66	0.67	0.68
	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
σ_η	0.056	0.056	0.056	0.056	0.056	0.056
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
σ_{ϵ^b}	0.23	0.22		0.29	0.27	
	(0.03)	(0.03)		(0.04)	(0.04)	
ρ	0.27			0.23		
	(0.12)			(0.14)		
$\rho_{\epsilon^a_{2-1}\epsilon^a_{3-1}}$	0.57	0.53	0.45			
	(0.10)	(0.10)	(0.10)			
$\sigma_{\epsilon^a_{3-1}}$	0.50	0.51	0.57	0.92	0.93	1.00
	(0.05)	(0.05)	(0.05)			
LL	111.9	109.4	94.9	88.5	86.4	71.8

Note: The number of draws is 10 for the GHK and 10 for ϵ^b_{jt}.

markets. Columns 4 through 6 present the results that are directly comparable (up to a constant) to the results of the logit case, i.e., V is the identity matrix multiplied by a constant. Columns 1 through 3 present the general case for V, where the independence of irrelevant alternatives property is relaxed, and the elements in the diagonal of V do not need to be all equal.

These results confirm the results obtained in the base case and show their robustness to allowing the independence of irrelevant alternatives not to hold. They show clearly the importance of considering the endogeneity of the pricing decision: (i) there are important common shocks across individuals (the null hypothesis $\sigma_{\epsilon^b} = 0$ is rejected), (ii) there is significant endogeneity of the price variable (the null hypothesis $\rho = 0$ is rejected), and (iii) the parameters that estimate the impact of price on choice are underestimated if endogeneity is not taken into account. For the yogurt market the price effect is underestimated by 65%; for the ketchup market, it is underestimated by 13% (in comparison to the case where endogeneity and common errors are not considered, i.e., column 1 versus column 3). Allowing for a more general error structure is also clearly important. The common demand shocks are on average 36% of the total demand shocks in the yogurt market and 19% in the ketchup market.

In order to check the impact of the number of draws on the previous results, we derived the results of the three left columns of Table 3 with a greater number of draws for the GHK (20) and for ϵ_{jt}^b (100). These results are presented in Table 4.

A comparison of the results of Table 4 to the first three columns of Table 3 suggests that the bias introduced by SML is not substantial. Furthermore, the SML method is underestimating the likelihood function when we relax the common errors and correlation restrictions, i.e., the SML method is underestimating the impact of endogeneity.

Finally, one can generalize the error structure even further by allowing η_{jt} and ϵ_{jt}^b to be distributed differently and correlated across the brands. In particular,

Table 4 Greater Number of Draws

	Complete Model	$\rho = 0$	$\rho = 0$, $\sigma_{\epsilon^b} = 0$
Yogurt Market			
Dannon	2.65	1.66	1.46
	(0.29)	(0.13)	(0.08)
Yoplait	2.77	1.76	1.52
	(0.31)	(0.15)	(0.10)
Price	−10.45	−7.06	−6.31
	(1.00)	(0.47)	(0.31)
Display	0.14	0.17	0.23
	(0.08)	(0.08)	(0.04)
Feature	0.004	−0.058	−0.018
	(0.04)	(0.04)	(0.03)
LPI	0.79	0.83	0.83
	(0.04)	(0.04)	(0.04)
σ_η	0.039	0.038	0.038
	(0.002)	(0.002)	(0.002)
σ_{ϵ^b}	0.29	0.19	
	(0.04)	(0.02)	
ρ	0.79		
	(0.06)		
$\rho_{\epsilon_2^a - 1 \epsilon_3^a - 1}$	0.69	0.71	0.68
	(0.03)	(0.03)	(0.03)
$\sigma_{\epsilon_3^a - 1}$	0.83	0.88	0.88
	(0.04)	(0.05)	(0.04)
LL	−1079.3	−1095.6	−1126.6
Ketchup Market			
DelMonte	−1.46	−1.34	−1.11
	(0.16)	(0.15)	(0.11)
Hunts	−1.45	−1.35	−1.16
	(0.15)	(0.13)	(0.10)
Price	−2.62	−2.37	−2.20
	(0.32)	(0.30)	(0.24)
Display	0.57	0.58	0.49
	(0.06)	(0.06)	(0.04)
Feature	0.39	0.34	0.42
	(0.09)	(0.09)	(0.05)
LPI	0.57	0.57	0.61
	(0.05)	(0.05)	(0.04)
σ_η	0.056	0.056	0.056
	(0.001)	(0.001)	(0.001)
σ_{ϵ^b}	0.26	0.25	
	(0.03)	(0.03)	
ρ	0.28		
	(0.14)		
$\rho_{\epsilon_2^a - 1 \epsilon_3^a - 1}$	0.64	0.60	0.46
	(0.09)	(0.10)	(0.10)
$\sigma_{\epsilon_3^a - 1}$	0.45	0.44	0.57
	(0.05)	(0.05)	(0.05)
LL	123.9	121.2	96.4

Note: The number of draws is 20 for the GHK and 100 for ϵ_{jt}^b.

for the cases we are analyzing (i.e., $J = 3$) we have the following parameters:

$$V_\eta = \mathrm{VAR}\begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{bmatrix} = \begin{bmatrix} \sigma^2_{\eta 1} & \rho_{\eta 12}\sigma_{\eta 1}\sigma_{\eta 2} & \rho_{\eta 13}\sigma_{\eta 1}\sigma_{\eta 3} \\ & \sigma^2_{\eta 2} & \rho_{\eta 23}\sigma_{\eta 2}\sigma_{\eta 3} \\ & & \sigma^2_{\eta 3} \end{bmatrix}, \tag{12}$$

$$V_{\epsilon^b} = \mathrm{VAR}\begin{bmatrix} \epsilon^b_{1t} \\ \epsilon^b_{2t} \\ \epsilon^b_{3t} \end{bmatrix} = \begin{bmatrix} \sigma^2_{\epsilon^b_1} & \rho_{\epsilon^b_{12}}\sigma_{\epsilon^b_1}\sigma_{\epsilon^b_2} & \rho_{\epsilon^b_{13}}\sigma_{\epsilon^b_1}\sigma_{\epsilon^b_3} \\ & \sigma^2_{\epsilon^b_2} & \rho_{\epsilon^b_{23}}\sigma_{\epsilon^b_2}\sigma_{\epsilon^b_3} \\ & & \sigma^2_{\epsilon^b_3} \end{bmatrix}, \tag{13}$$

$$V_{\epsilon^b\eta} = \mathrm{COV}\begin{bmatrix} \epsilon^b_{1t} \\ \epsilon^b_{2t} \\ \epsilon^b_{3t} \end{bmatrix}\begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{bmatrix}$$
$$= \begin{bmatrix} \rho_{\epsilon^b_1\eta 1}\sigma_{\epsilon^b_1}\sigma_{\eta 1} & \rho_{\epsilon^b_1\eta 2}\sigma_{\epsilon^b_1}\sigma_{\eta 2} & \rho_{\epsilon^b_1\eta 3}\sigma_{\epsilon^b_1}\sigma_{\eta 3} \\ \rho_{\epsilon^b_2\eta 1}\sigma_{\epsilon^b_2}\sigma_{\eta 1} & \rho_{\epsilon^b_2\eta 2}\sigma_{\epsilon^b_2}\sigma_{\eta 2} & \rho_{\epsilon^b_2\eta 3}\sigma_{\epsilon^b_2}\sigma_{\eta 3} \\ \rho_{\epsilon^b_3\eta 1}\sigma_{\epsilon^b_3}\sigma_{\eta 1} & \rho_{\epsilon^b_3\eta 2}\sigma_{\epsilon^b_3}\sigma_{\eta 2} & \rho_{\epsilon^b_3\eta 3}\sigma_{\epsilon^b_3}\sigma_{\eta 3} \end{bmatrix}. \tag{14}$$

Now, because we are only able to measure the difference between latent utilities (because the choice probabilities depend only on the differences between latent utilities), these parameters are underidentified (because they define the latent utilities for all the brands). We thus have to impose 6 restrictions on the above parameters. In order to make comparisons with the previous results we chose to impose the following restrictions: $\rho_{\epsilon^b_{13}} = 0$, $\rho_{\epsilon^b_{23}} = 0$, $\sigma_{\epsilon^b_3} = \sigma_{\epsilon^b_1}$, $\rho_{\epsilon^b_3}\eta_1 = 0$, $\rho_{\epsilon^b_3}\eta_2 = 0$, and $\rho_{\epsilon^b_3}\eta_3 = \rho_{\epsilon^b_1}\eta_3$. Similarly to any identifying restrictions related to only observing relative preference among brands, these do not affect the implications of the model, i.e., the exact magnitude of the impact of the exogenous variables on the choice among brands.

The log-likelihood function is now changed to

$$L(\beta, \alpha_0, \alpha_1, V_{\epsilon^b}, V_\eta, V_{\epsilon^b\eta})$$
$$= \prod_{t=1}^{T} f(\eta_{ht}\ \forall\, h;\ V_\eta;\ \alpha_0, \alpha_1) g(d_{it}\ \forall\, i \in I_t | \eta_{jt}\ \forall\, j) \tag{15}$$

where $f(\eta_{ht}\ \forall h;\ V_\eta;\ \alpha_0, \alpha_1)$ is the density function of a normal random vector with mean zero and variance V_η, and $g(d_{it}\ \forall i \in I_t|\eta_{jt}\ \forall j)$ is the probability of observing choices $d_{it}\ \forall i \in I_t$ in week t given $\eta_{jt}\ \forall j \in J$. Finally, $g(d_{it}\ \forall i \in I_t|\eta_{jt}\ \forall j)$ can be written as

$$g(d_{it}\ \forall\, i \in I_t | \eta_{jt}\ \forall\, j)$$
$$= \int\int \cdots \int \prod_{i \in I_t} \mathrm{Prob}(j_{it};\ i,\ t,\ \epsilon^b_{jt}\ \forall\, j,\ t)$$
$$\cdot f(\epsilon^b_{kt}\ \forall\, k | \eta_{kt}\ \forall\, k;\ V_{\epsilon^b\eta}, V_{\epsilon^b}, V_\eta) \prod_{k=1}^{J} d\epsilon^b_{kt} \tag{16}$$

where $f(\epsilon^b_{kt}\ \forall k|\eta_{kt}\ \forall k;\ V_{\epsilon^b\eta}, V_{\epsilon^b}, V_\eta)$ is the density of the vector ϵ^b_t given η_t.

Testing for endogeneity is testing for $V_{\epsilon^b\eta} = 0$. Testing for common demand shocks across consumers is testing for $V_{\epsilon^b} = 0$. The results are presented in Tables 5 and 6.

These results basically confirm the results obtained above: common errors in the choice model and endogeneity in the marketing-mix variables (i.e., price) matter substantially. It is interesting in this model to interpret the correlations $\rho_{\epsilon^b_j}\eta_{j'}$. They can be interpreted as the direction of the price reaction of seller j' to a positive common shock to the demand of product j (given the identifying constraints being used). In this sense, note that the yogurt market results indicate that a positive shock in the demand for Dannon causes the prices of Yoplait and the Private Label to increase, but a positive shock on the demand for Yoplait causes the prices of Dannon and the Private Label to fall (given the identifying constraints). In the ketchup market any positive shock in a demand of a product causes the prices of all products to increase. Note also that in both markets all $\rho_{\epsilon^b_j}\eta_j$ are always positive, i.e., as expected, a positive shock on the demand of a product causes the seller of that product to increase its price.

Furthermore, the results in Tables 5 and 6 show that the constraints that were relaxed in this generalization (allowing for richer V_η, V_{ϵ^b}, and $V_{\epsilon^b\eta}$) are rejected, i.e., it is important to allow for rich error

Table 5 Generalized Model: Yogurt Market

	Complete Model	Zero Correlations	Zero Common Errors
Dannon	2.35	1.62	1.49
	(0.25)	(0.13)	(0.09)
Yoplait	2.48	1.73	1.55
	(0.26)	(0.14)	(0.10)
Price	−9.46	−6.92	−6.40
	(0.88)	(0.47)	(0.35)
Display	0.21	0.22	0.24
	(0.09)	(0.08)	(0.05)
Feature	0.018	−0.030	−0.016
	(0.05)	(0.04)	(0.03)
LPI	0.86	0.84	0.85
	(0.05)	(0.04)	(0.04)
$\sigma_{\eta 1}$	0.053	0.053	0.053
	(0.005)	(0.006)	(0.006)
$\rho_{\eta 12}$	0.0019	0.0078	0.0078
	(0.23)	(0.20)	(0.20)
$\rho_{\eta 13}$	0.072	0.030	0.030
	(0.14)	(0.14)	(0.14)
$\sigma_{\eta 2}$	0.024	0.024	0.024
	(0.001)	(0.001)	(0.001)
$\rho_{\eta 23}$	0.056	0.040	0.040
	(0.16)	(0.16)	(0.16)
$\sigma_{\eta 3}$	0.031	0.031	0.031
	(0.003)	(0.002)	(0.002)
$\sigma_{\epsilon_1^b}$	0.20	0.14	
	(0.04)	(0.03)	
$\rho_{\epsilon_{12}^b}$	−0.20	0.18	
	(0.29)	(0.15)	
$\sigma_{\epsilon_2^b}$	0.26	0.22	
	(0.05)	(0.05)	
$\rho_{\epsilon_1^b \eta 1}$	0.66		
	(0.17)		
$\rho_{\epsilon_1^b \eta 2}$	0.030		
	(0.39)		
$\rho_{\epsilon_1^b \eta 3}$	0.24		
	(0.28)		
$\rho_{\epsilon_2^b \eta 1}$	−0.36		
	(0.16)		
$\rho_{\epsilon_2^b \eta 2}$	0.62		
	(0.19)		
$\rho_{\epsilon_2^b \eta 3}$	−0.015		
	(0.21)		
$\rho_{\epsilon_{2-1}^a \epsilon_{3-1}^a}$	0.70	0.71	0.67
	(0.04)	(0.03)	(0.03)
$\sigma_{\epsilon_{3-1}^a}$	0.93	0.91	0.90
	(0.06)	(0.05)	(0.05)
LL	−1039.6	−1061.0	−1089.0

Note: The number of draws is 10 for the GHK and 10 for ϵ_{jt}^b.

Table 6 Generalized Model: Ketchup Market

	Complete Model	Zero Correlations	Zero Common Errors
DelMonte	−1.31	−1.14	−1.10
	(0.17)	(0.14)	(0.11)
Hunts	−1.32	−1.16	−1.16
	(0.16)	(0.13)	(0.11)
Price	−2.50	−2.17	−2.19
	(0.38)	(0.30)	(0.25)
Display	0.43	0.47	0.49
	(0.07)	(0.05)	(0.04)
Feature	0.47	0.35	0.42
	(0.11)	(0.08)	(0.05)
LPI	0.56	0.57	0.60
	(0.04)	(0.05)	(0.05)
$\sigma_{\eta 1}$	0.029	0.029	0.029
	(0.002)	(0.001)	(0.001)
$\rho_{\eta 12}$	0.038	0.019	0.019
	(0.24)	(0.24)	(0.24)
$\rho_{\eta 13}$	−0.027	−0.062	−0.062
	(0.13)	(0.11)	(0.11)
$\sigma_{\eta 2}$	0.081	0.081	0.081
	(0.004)	(0.005)	(0.005)
$\rho_{\eta 23}$	0.025	0.016	0.016
	(0.18)	(0.19)	(0.19)
$\sigma_{\eta 3}$	0.044	0.044	0.044
	(0.002)	(0.002)	(0.002)
$\sigma_{\epsilon_1^b}$	0.20	0.19	
	(0.03)	(0.03)	
$\rho_{\epsilon_{12}^b}$	−0.24	−0.97	
	(0.17)	(0.14)	
$\sigma_{\epsilon_2^b}$	0.30	0.21	
	(0.10)	(0.08)	
$\rho_{\epsilon_1^b \eta 1}$	0.64		
	(0.10)		
$\rho_{\epsilon_1^b \eta 2}$	0.38		
	(0.17)		
$\rho_{\epsilon_1^b \eta 3}$	0.54		
	(0.17)		
$\rho_{\epsilon_2^b \eta 1}$	0.44		
	(0.21)		
$\rho_{\epsilon_2^b \eta 2}$	0.51		
	(0.19)		
$\rho_{\epsilon_2^b \eta 3}$	0.40		
	(0.22)		
$\rho_{\epsilon_{2-1}^a \epsilon_{3-1}^a}$	0.62	0.53	0.45
	(0.10)	(0.11)	(0.11)
$\sigma_{\epsilon_{3-1}^a}$	0.47	0.49	0.57
	(0.05)	(0.05)	(0.05)
LL	188.9	178.6	161.6

Note: The number of draws is 10 for the GHK and 10 for ϵ_{jt}^b.

structures. However, the parameters of the choice model do not change substantially.[24]

4. Conclusion

Our results show the importance of accounting for endogeneity in brand choice models estimated using scanner panel data from two product categories. Variables which are unobserved to the marketing researchers seem to be used by marketing managers to set the levels of the decision variables and to affect consumer brand choice. The results in this paper show that not accounting for endogeneity may result in underestimation of the market response to price. Using lagged prices as instruments, the size of the bias was larger for the yogurt market than the ketchup market. However, this difference disappeared when we used better (but harder to find) instruments such as cost variables. The use of these instruments also showed that the lagged price instruments may be underestimating the size of the bias. We also showed that these results are robust to having a loyalty variable with exponential smoothing of past purchases, and to more complex error structures.

We also tested the robustness of the results presented here to the case where the no-purchase decision is in the model (the unconditional model). The results, available upon request from the authors, show that there is endogeneity both in the purchase and the choice decision, and the results presented here are substantially strengthened.

There are several limitations to this study. First, we tested for endogeneity in only two product categories with a limited number of alternatives. Tests for endogeneity in other product categories allowing for a greater number of alternatives are still needed. Second, we only analyzed endogeneity in the price variable. Several other explanatory variables are also probably affected by endogeneity, for example, displays and features. Third, several aspects of the execution of our tests could be improved upon with greater computer capabilities: greater number of simulations, allowing for both additive and slope endogeneity at the same time, allowing for more general heterogeneity structures, including in the model both very general error structures, and heterogeneity, accounting for different seasonal effects across brands. Fourth, one could check the robustness of the results to other estimation methods, in particular, to estimation methods that rely on less distribution assumptions. Finally, recent developments in the literature on brand choice such as the when/what/how much decision (Gupta 1988, Chiang 1991, Chintagunta 1993) could also be included in the model.

Nevertheless, we believe that our results may be robust to all these extensions of the base model: Endogeneity in the marketing-mix variables may be important and can create bias in the parameter estimates in several markets.[25,26]

[24] Another potentially important generalization of the error structures, not studied here, is to allow the common shocks ϵ_{jt}^{b} to be autocorrelated through time. This specification can be particularly important because shifts in preferences for brands are not likely to be i.i.d., but in fact strongly correlated through time. With the current computational capabilities and this likelihood approach, this structure is however relatively complicated to estimate because one has both cross-section and time-series correlation of the error terms. Note, however, that if ϵ_{jt}^{b} is stationary and ergodic we still have consistent estimators of the parameters of the utility equation with the estimation method presented above.

[25] One easily applicable test for endogeneity is the test of the existence of common shocks in the utility equation, which might be a strong cue for the existence of endogeneity problems. The test is a lagrange multiplier test on the restriction that the variance of the common shocks is zero. Details are available in Newey and McFadden (1994).

[26] We are grateful to Daniel McFadden for long and insightful conversations. We are also grateful to Tülin Erdem, Michael Keane, Don Lehmann, Ram Rao, Pablo Spiller, and Naufel Vilcassim for comments on an earlier version of this paper. All remaining errors are our responsibility alone. This work is supported by NSF grant SES-9309812. Research assistance by Mark Stiving and Ying Zhao is gratefully appreciated.

References

Bass, F. 1969. A simultaneous equation regression study of advertising and sales of cigarettes. *J. Marketing Res.* **6** 291–300.

Berry, S. 1994. Estimating discrete-choice models of product differentiation. *Rand J. Econom.* **25** 242–262.

——, J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63** 841–890.

Besanko, D., S. Gupta, D. Jain. 1998. Logit demand estimation under competitive pricing behavior: an equilibrium framework. *Management Sci.* **44** 1533–1547.

Bresnahan, T., S. Stern, M. Trajtenberg. 1997. Market segmentation and the sources of rents from innovation: personal computers in the late 1980s. *Rand J. Econom.* **28** S17–S44.

Chiang, J. 1991. A simultaneous approach to the whether, what, and how much to buy questions. *Marketing Sci.* **10** 297–315.

Chintagunta, P. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Sci.* **12** 184–208.

——, D. Jain. 1995. Empirical analysis of a dynamic duopoly model of competition. *J. Econom. Management Strategy* **4** 109–131.

Fader, P., J. Lattin, J. Little. 1992. Estimating nonlinear parameters in the multinomial logit model. *Marketing Sci.* **11** 372–385.

Gonul, F., K. Srinivasan. 1993. Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues. *Marketing Sci.* **12** 213–229.

Gourieroux, C., A. Monfort. 1993. Simulation-based inference: a survey with special reference to panel data models. *J. Econometrics* **59** 5–33.

Green, W. 1997. *Econometric Analysis*, 3rd ed. Prentice-Hall, Englewood Cliffs, NJ.

Guadagni, P., J. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2** 203–238.

Gupta, S. 1988. Impact of sales promotion on when, what, and how much to buy. *J. Marketing Res.* **25** 342–355.

Hajivassiliou, V., D. McFadden. 1993. The method of simulated scores for the estimation of LDV models with an application to external debt crisis. Mimeo. University of California, Berkeley, CA.

——, ——, P. Ruud. 1992. Simulation of multivariate normal orthant probabilities: theoretical and computational results. Mimeo. University of California, Berkeley, CA.

Heckman, J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* **46** 931–959.

——. 1979. Sample selection bias as a specification error. *Econometrica* **47** 153–161.

Kadiyali, V., N. Vilcassim, P. Chintagunta. 1995. Empirical analysis of competitive product line pricing decisions: Lead, follow, or move together? *J. Bus.* **69** 459–487.

Keane, M. 1993. Simulation estimation for panel data models with limited dependent variables. C. R. Rao, G. S. Maddala, H. D. Vinod, eds., *The Handbook of Statistics* 11: *Econometrics*. Elsevier Science Publishers, Amsterdam, 545–572.

McFadden, D. 1973. Conditional choice analysis of qualitative choice behavior. P. Zarembka, ed., *Frontiers of Econometrics*, Academic Press, New York, 105–142.

Milgrom, P., J. Roberts. 1986. Prices and advertising signals of product quality. *J. Political Econom.* **94** 796–821.

Morgan, M. 1990. *The History of Econometric Ideas*. Cambridge University Press, New York.

Newey, W. 1985. Maximum likelihood estimation testing and conditional moment tests. *Econometrica* **53** 1047–1070.

——. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *J. Econometrics* **36** 230–251.

——, D. McFadden. 1994. Large sample estimation and hypothesis testing. Z. Griliches, M. Intriligator, eds. *Handbook of Econometrics* 4. Elsevier Science Publishers, Amsterdam.

Rivers, D., Q. H. Vuong. 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *J. Econometrics* **39** 347–366.

Tellis, G. J. 1988. The price elasticity of selective demand: a meta-analysis of econometric models of sales. *J. Marketing Res.* **25** 331–341.

——, F. Zufryden. 1995. Tackling the retailer decision maze: which brands to discount, how much, when, and why? *Marketing Sci.* **14** 271–299.

Villas-Boas, J. M., R. Winer. 1994. Endogeneity in brand choice models. Mimeo. University of California, Berkeley, CA.

Winer, R. 1986. A reference price model of brand choice for frequently purchased products. *J. Consumer Res.* **13** 250–256.

Wooldridge, J. 1994. Estimation and inference for dependent processes. R. Engle, D. McFadden, eds. *Handbook of Econometrics* 4. Elsevier Science Publishers, Amsterdam, 2639–2738.

Accepted by Dipak C. Jain; received May 1998. This paper has been with the authors 11 months for 3 revisions.