A Control Function Approach to Endogeneity in Consumer Choice Models

Author(s): Amil Petrin and Kenneth Train

Source: *Journal of Marketing Research*, Feb., 2010, Vol. 47, No. 1 (Feb., 2010), pp. 3-13

Published by: Sage Publications, Inc. on behalf of American Marketing Association

Stable URL: https://www.jstor.org/stable/20618950

AMIL PETRIN and KENNETH TRAIN*

Endogeneity arises for numerous reasons in models of consumer choice. It leads to inconsistency with standard estimation methods that maintain independence between the model's error and the included variables. The authors describe a control function approach for handling endogeneity in choice models. Observed variables and economic theory are used to derive controls for the dependence between the endogenous variable and the demand error. The theory points to the relationships that contain information on the unobserved demand factor, such as the pricing equation and the advertising equation. The authors' approach is an alternative to Berry, Levinsohn, and Pakes's (1995) product-market controls for unobserved quality. The authors apply both methods to examine households' choices among television options, including basic and premium cable packages, in which unobserved attributes, such as quality of programming, are expected to be correlated with price. Without correcting for endogeneity, aggregate demand is estimated to be upward-sloping, suggesting that omitted attributes are positively correlated with demand. Both the control function method and the product-market controls method produce downward-sloping demand estimates that are similar.

Keywords: customer choice, endogeneity, advertising, price effects, econometric models

A Control Function Approach to Endogeneity in Consumer Choice Models

There are several discrete choice demand settings in which researchers have shown that factors not included in the analysis are correlated with the included factors, thus violating the standard independence assumption for consistency (see, e.g., Bass 1969; Berry 1994). In these cases, the estimated impact of the observed factor on demand captures not only that factor's effect but also the effect of the unobserved factors that are correlated with it. For example, products with higher quality usually have higher prices both because the attributes are costly to provide and because they raise demand. When some product attributes are either not observed by the researcher or difficult to measure, such as stylishness of design, estimated price elasticities will be biased in the positive direction.[1]

The problem is often exacerbated by the difficulty of signing this bias. Consider estimated price elasticities with unobserved advertising. Optimizing firms maximize profits with respect to both price and advertising, so in general, they cannot be independent. Firms might raise the price of their products when they advertise if they believe that it stimulates demand. Alternatively, firms may lower price when they advertise (e.g., as a part of a sale). The possibility of either case makes the sign of the bias ambiguous.[2]

*Amil Petrin is Associate Professor of Economics, Department of Economics, University of Minnesota (e-mail: petrin@umn.edu). Kenneth Train is Adjunct Professor of Economics and Public Policy, Department of Economics, University of California, Berkeley (e-mail: train@econ.berkeley.edu). The authors thank the two anonymous *JMR* reviewers for their useful comments and gratefully acknowledge support from the National Bureau of Economic Research. Pradeep Chintagunta served as associate editor for this article.

[1] Another common bias arises when estimating the elasticity of demand for recreation, shopping, theater, and so on, with respect to travel time. If people with strong tastes for such activities live close to these activities, travel time to the desired activity will be negatively correlated with unobserved taste, leading to a negative bias in the elasticity of demand with respect to travel time.

[2] Variation in unobserved variables, including unmeasured attributes and advertising, may itself be caused by changes in demand conditions, such as shifts in tastes. In this case, a model that describes the variation in demand and its relationship to the unobserved variables would more fully represent the situation.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

In this article, we propose a control function method for alleviating bias in discrete choice demand settings.[3] The approach includes extra variables in the empirical specification to condition out the variation in the unobserved factor that is not independent of the endogenous variable. We derive these controls using economic theory to point to alternative equations that contain information about the unobserved demand factor. Although our empirical application focuses on the pricing equation, any equation that contains information on relevant unobserved factors may be available for use. For example, we anticipate that researchers will explore the advertising equation, which is also affected by the unobserved demand factor.

The most widely used bias correction method in discrete choice demand settings is the product-market control approach developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995; hereinafter, BLP) for market-level data and then extended to consumer-level data (Berry, Levinsohn, and Pakes 2004; Goolsbee and Petrin 2004). The approach appeals to the aggregate demand equations as a source of information on the unobserved demand factor and has been applied to consumers' choice among television options (Crawford 2000; Goolsbee and Petrin 2004), minivans (Petrin 2002), and grocery goods (Chintagunta, Dubé, and Goh 2005; Nevo 2001), to name only a few.

Our control function approach provides a useful alternative to the BLP approach. The control function approach is both easier to estimate and available in some situations in which the BLP estimator is not valid. For example, the BLP approach is not consistent in settings in which there are zero, one, or just a small number of purchase observations per product, because it requires that market shares be observed with relatively little sampling error (see Berry, Linton, and Pakes 2004). The BLP approach is also not available for many recently developed empirical demand models, which either maintain assumptions that are not consistent with the BLP setting or are sufficiently complicated to preclude estimating the BLP controls (e.g., Bajari et al. 2007; Fox 2008; Hendel and Nevo 2006). In contrast, our control function approach simply adds new regressors to the demand specification, making it available in all these settings.

Either approach is applicable in our empirical application, so we estimate both for comparison. We also provide more discussion relating the approaches in the section "Comparison with Product-Market Control."

Other methods related to endogeneity in demand settings have been developed. Louviere and colleagues (2005) describe the various manifestations of endogeneity in marketing contexts and the implications for estimation. Building on work by Villas-Boas and Winer (1999), Kuksov and Villas-Boas (2008) describe methods for testing for endogeneity. Villas-Boas and Winer (1999) and Gupta and Park (2009) have developed a maximum likelihood approach, and Yang, Chen, and Allenby (2003) and Jiang, Manchanda, and Rossi (2007) have developed Bayesian methods for handling endogeneity.

In the following sections, we describe the control function approach, provide example specifications, discuss the relationship to pricing behavior, and illustrate our approach with an application to households' choices among television options.

MODEL

Consumer n chooses one of the J competing alternatives. The utility that the consumer obtains from alternative j is as follows:

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}, \tag{1}$$

where y_{nj} is the observed endogenous variable, x_{nj} is a vector of observed exogenous variables that affect the utility derived from choice j, β_n are parameters that represent the tastes of consumer n, and ε_{nj} is the unobserved utility.[4] The endogenous variable might be price, advertising, travel time, or whatever is relevant in the context. The econometric problem arises because ε_{nj} is not independent of y_{nj}, as maintained by standard estimation techniques.

The idea behind the control function correction is to derive a proxy variable that conditions on the part of y_{nj} that depends on ε_{nj}. If this can be done, the remaining variation in the endogenous variable will be independent of the error, and standard estimation approaches will again be consistent.

In this discrete choice context, the approach posits that y_{nj} can be written as a function of all exogenous variables entering utility for any of the choices, denoted as x_n; the variables z_n that do not enter utility directly but affect y_{nj} (typically the instruments); and a vector of J unobserved terms, μ_n:

$$y_{nj} = W(x_n, z_n, \mu_n). \tag{2}$$

The approach maintains that μ_n and ε_{nj} are independent of x_n and z_n but are not independent of each other. This equation illustrates the source of the dependence between y_{nj} and ε_{nj}—that is, μ_n affects y_{nj} and is also not independent of ε_{nj}.

The key to the control function approach is to note that under the maintained assumptions, conditional on μ_n, ε_{nj} is independent of y_{nj}. The feasibility of the control function approach in any setting will be determined by whether the practitioner is able to recover μ_n so it can be conditioned on when the parameters are estimated.[5]

We analyze the control function case when y_{nj} is additive in its observed and unobserved covariates. A special, illustrative case is when there is a single unobserved factor μ_{nj} for each choice j:

$$y_{nj} = W(x_n, z_n; \gamma) + \mu_{nj}, \tag{3}$$

[3]Heckman and Robb (1985) introduced the term "control function" in the context of selection models, but the concepts date back at least to Heckman (1978) and Hausman (1978). The method has been applied to a Tobit model by Smith and Blundell (1986) and to binary probit models by Rivers and Vuong (1988). Blundell and Powell (2004) include it in their discussion of semiparametric methods for binary choice.

[4]"Observed" and "unobserved" are from the perspective of the practitioner. All terms are observed by the consumer when making the decision.

[5]As Imbens and Newey (2008) show, it is sufficient to condition on any one-to-one function of μ_n.

where we make explicit γ the parameters of this function. With additivity and the independence assumptions, the controls μ_{nj} are straightforward to recover using any standard estimator (e.g., ordinary least squares). The question becomes how the new controls are entered into the utility function to condition out the dependence between y_{nj} and ε_{nj}.

One approach enters μ_{nj} in a flexible manner to condition out any function of it. Decomposing ε_{nj} into the part that can be explained by a general function of μ_{nj} and the residual yields the following:

$$\varepsilon_{nj} = CF(\mu_{nj}; \lambda) + \tilde{\varepsilon}_{nj}, \tag{4}$$

where $CF(\mu_{nj}; \lambda)$ denotes the control function with parameters λ. The simplest approximation is to specify the control function as linear in μ_{nj}, in which case the control function is $CF(\mu_{nj}; \lambda) = \lambda\mu_{nj}$, λ is a scalar, and utility is as follows:

$$U_{nj} = V(y_{nj}, x_n, \beta_n) + \lambda\mu_{nj} + \tilde{\varepsilon}_{nj}. \tag{5}$$

Alternatively, we could allow for a polynomial approximation, adding higher-order terms of μ_{nj} and the necessary additional parameters.

More generally, it is possible to condition on the entire vector of controls μ_n for any choice j when calculating the control function. In this case, we have the following:

$$\varepsilon_{nj} = CF(\mu_n; \lambda) + \tilde{\varepsilon}_{nj}, \tag{6}$$

which can be approximated to first order with a vector of parameters $CF(\mu_n; \lambda) = \lambda'\mu_n$.

Again, higher-order terms are straightforward to add, though parameters increase rapidly in the number of alternative choices. Given the researcher's chosen control function specification, we then have the following:

$$U_{nj} = V(y_{nj}, x_n, \beta_n) + CF(\mu_n, \lambda) + \tilde{\varepsilon}_{nj}. \tag{7}$$

Conditional on μ_n, the probability that consumer n chooses alternative i is equal to

$$P_{ni} = \int I(U_{ni} > U_{nj}\ \forall j \neq i) f(\beta_n, \tilde{\varepsilon}_n) d\beta_n d\tilde{\varepsilon}_n, \tag{8}$$

where $f(\cdot)$ is the joint density of β_n and $\tilde{\varepsilon}_n$ and $I(\cdot)$ is the indicator function. All that remains to complete the specification is a distributional assumption applied to $f(\cdot)$.

The usual approach is to choose specific functional forms for the distribution of β_n and ε_{nj} (e.g., normal, logit), though these are typically difficult to motivate with economic theory. They are almost always chosen to be independent of each other, and we maintain that assumption here, which implies in our setting that β_n and $\tilde{\varepsilon}_{nj}$ are independent because conditioning on μ_n cannot induce dependence. In our application, we use normal and logit, though researchers can use whatever assumptions they desire to suit their setup. As in any application, checking the robustness to distributional assumptions is important.

The model is estimated in two steps. First, the endogenous variable is regressed on observed choice characteristics and the instruments. The residuals of this regression are retained and used to calculate the control function. Second, the choice model is estimated with the control function entering as an extra variable or variables.

Because the second step uses an estimate of μ_n from the first step, as opposed to the true μ_n, the asymptotic sampling variance of the second-step estimator needs to take this extra source of variation into account. Either the bootstrap can be implemented, or the standard formulas for two-step estimators can be used (Murphy and Topel 1985; Newey and McFadden 1994). Karaca-Mandic and Train (2003) derive the specific form of these formulas that is applicable to the control function approach. As they note, the bootstrap and asymptotic formulas provide similar standard errors for the application that we describe in our empirical results.

PARAMETRIC FUNCTIONAL FORMS

We consider several parametric forms for the errors in both equations. These parametric forms lead to direct parametric forms for the control function itself and the distribution of the demand residuals conditional on the controls. Although they need not be maintained, they provide an alternative to entering μ_n flexibly in utility and then choosing a distributional assumption for $\tilde{\varepsilon}_{nj}$.

Example 1: Jointly Normal Errors, Independent Over j

Suppose that μ_{nj} and ε_{nj} are jointly normal for each j and i.i.d. over j. Then,

$$CF(\mu_n; \lambda) = E(\varepsilon_{nj}|\mu_n) = \lambda\mu_{nj} \tag{9}$$

for each j and the deviations $\tilde{\varepsilon}_{nj} = \varepsilon_{nj} - CF(\mu_n; \lambda)$ are independent of μ_{nj} and all other regressors. Thus, the control function for each alternative is the residual from the endogenous variable regression interacted with λ, the one coefficient to be estimated. Utility is as follows:

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \tilde{\varepsilon}_{nj}, \tag{10}$$

where $\tilde{\varepsilon}_{nj}$ is i.i.d. normal with zero mean.

If β_n is fixed, the model is an independent probit with the residual entering as an extra variable. If β_n is random, the model is a mixed independent probit, mixed over the density of β_n (Train 2003, chaps. 5 and 6 on probit and mixed logit). However, note that the scale of the estimated model differs from that of the original model. In particular, $Var(\tilde{\varepsilon}_{nj}) < Var(\varepsilon_{nj})$, such that normalizing by setting $Var(\tilde{\varepsilon}_{nj}) = 1$ raises the magnitude of coefficients relative to the normalization $Var(\varepsilon_{nj}) = 1$.

Example 2: Extreme Value and Joint Normal Error Components, Independent over j

The previous example can be modified to generate a mixed logit, which has the same normalization for scale in the original and estimated model. This is one of the specifications that Villas-Boas and Winer (1999, Section 2) use. Let $\varepsilon_{nj} = \varepsilon^1_{nj} + \varepsilon^2_{nj}$, where ε^1_{nj} and μ_{nj} are jointly normal and ε^2_{nj} is i.i.d. extreme value for all j. Then, utility with the control function is

$$U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \sigma\eta_{nj} + \varepsilon^2_{nj}, \tag{11}$$

where η_{nj} is i.i.d. standard normal. The model is a mixed logit, with mixing over the error components η_{nj}, whose standard deviation σ is estimated, as well as over the random elements of β_n. The scale in the original utility is normalized by setting the scale of the extreme value distribution for ε^2_{nj}.

Example 3: Extreme Value and Joint Normal Error Components, Correlation over j

The generalization is straightforward conceptually but increases the number of parameters considerably. Let ε^1_n and μ_n be jointly normal with zero mean and covariance Ω. This covariance matrix is $2J \times 2J$ and is composed of submatrices labeled $\Omega_{\mu\mu}$, $\Omega_{\mu\varepsilon}$, $\Omega_{\varepsilon\varepsilon}$. Then, $CF(\mu_n; \lambda) = E(\varepsilon^1_n|\mu_n) = \Lambda\mu_n$, where the elements of matrix Λ are related to the elements of Ω: $\Lambda = \Omega_{\mu\varepsilon}\Omega^{-1}_{\mu\mu}$. Stacked utilities then become

$$U_n = V(y_n, x_n, \beta_n) + \Lambda\mu_n + \Gamma\eta_n + \varepsilon^2_n, \tag{12}$$

where η_n is now a vector of J i.i.d. standard normal deviates and Γ is the lower-triangular Choleski matrix of $\Omega_{\varepsilon\varepsilon} - \Omega_{\mu\varepsilon}\Omega^{-1}_{\mu\mu}\Omega_{\mu\varepsilon}$. In this case, the residuals for each alternative enter the utility of all alternatives, and the mixing is over a set of J normal error components. Villas-Boas and Winer (1999, Section 3) generalize this specification further by allowing ε^2_n to be correlated over alternatives, specifying it to be normally distributed instead of extreme value to accommodate this correlation.

PRICING BEHAVIOR AND THE CONTROL FUNCTION APPROACH

Consider consumers' choice among products, where the endogenous variable y_{nj} is price p_{nj}. We investigate the control function approach using some variant of the controls suggested previously with both marginal cost and monopoly pricing. The utility that consumer n obtains from product j is specified as in Example 2:

$$U_{nj} = V(p_{nj}, x_{nj}, \beta_n) + \varepsilon^1_{nj} + \varepsilon^2_{nj}, \tag{13}$$

where ε^1_{nj} is correlated with price and ε^2_{nj} is i.i.d. extreme value. Here, ε^1_{nj} might represent unobserved attributes of the product that are not independent of price. Typically, prices vary over people because different people are in different markets.

The marginal cost of product j in consumer n's choice set is denoted as $MC(z_{nj}, \nu_{nj})$, where z_{nj} are exogenous variables observed by the analyst and ν_{nj} is unobserved. The observed variables z_{nj} will typically overlap with x_{nj} insofar as observed attributes of the product affect both demand and cost.

Marginal Cost Pricing

Consider marginal cost (MC) pricing, and assume that MC is separable in the unobserved term. The pricing equation becomes

$$p_{nj} = MC(z_{nj}, \nu_{nj}) = W(z_{nj}, \gamma) + \nu_{nj}, \tag{14}$$

where γ are parameters to be estimated. Following Example 2, assume that ε^1_{nj} and ν_{nj} are jointly normal, i.i.d. over j. Correlation may arise, for example, because unobserved attributes affect utility as well as costs, thus entering both ε^1_{nj} and ν_{nj}. Given the separability assumption on the unobserved term in the pricing equation, utility becomes

$$U_{nj} = V(p_{nj}, x_{nj}, \beta_n) + \lambda\nu_{nj} + \sigma\eta_{nj} + \varepsilon^2_{nj}, \tag{15}$$

where η_{nj} is i.i.d. standard normal. The same specification is appropriate when there is a constant markup over cost in the determination of prices.

Monopoly Pricing

Consider monopoly pricing, where price depends on the elasticity of demand and marginal cost. The pricing equation for a monopolist is as follows:[6]

$$p_{nj} = [p_{nj}/|e(\varepsilon^1_n)|] + MC(z_{nj}, \nu_{nj}), \tag{16}$$

where $e(\varepsilon^1_n)$ is the elasticity of demand. This elasticity depends on all factors that affect demand, including attributes that are not observed by the analyst. The elasticity is written as a function of ε^1_n to explicitly denote this dependence.

Now, suppose that the analyst estimates the following price equation:

$$p_{nj} = W(x_{nj}, z_{nj}, \gamma) + \mu_{nj}. \tag{17}$$

Then, ε^1_n enters the pricing equation in a nonseparable manner, suggesting that the additively separable μ_{nj} will not fully condition out the entire dependence of p_{nj} on ε^1_{nj}. Any remaining dependence will bias the estimated price elasticitiy.

In this situation, Villas-Boas (2007) suggests working in the reverse direction. Instead of specifying the joint distribution of $\langle\nu_n, \varepsilon^1_n\rangle$ and then deriving the implications for the distribution of $\langle\mu_n, \varepsilon^1_n\rangle$, Villas-Boas shows that if the price of each product is strictly monotonic in its marginal cost, there is a distribution of $\langle\nu_n, \varepsilon^1_n\rangle$ and a marginal cost function that is consistent with any given distribution of $\langle\mu_n, \varepsilon^1_n\rangle$. This result implies that the analyst can specify a distribution of ε^1_n conditional on μ_n, as needed for the control function approach, and knows that there is some distribution of ν_n and ε^1_n that gives rise to it.

EMPIRICAL MODEL AND DATA

To illustrate, we apply the control function approach to households' choice of television reception options. The specification and data are similar to those of Goolsbee and Petrin (2004), who apply the BLP approach. By using a situation in which both approaches can be applied, we are able to compare results.

In general, households have four alternatives for television: (1) antenna-only, (2) cable with basic or extended service, (3) cable with a premium service added (e.g., HBO), and (4) satellite dish. Basic and extended cable are

[6]The more common form of this equation is $(p-MC)/p = 1/|e|$.

combined because the data do not differentiate which of these options the households chose. Goolsbee and Petrin (2004) describe the market for cable and satellite television, emphasizing the importance of accounting for endogeneity of price, which arises because unobserved attributes of cable television, like the quality of programming, are not independent of price.

Our sample consists of 11,810 households in 172 geographically distinct markets. Each market contains one cable franchise that offers basic, extended, and premium packages. There are several multiple-system operators, such as AT&T and Time Warner, which own many cable franchises throughout the country (thus serving several markets). The price and other attributes of the cable options vary over markets, even for markets served by the same multiple-system operator. Satellite prices do not vary geographically, and the price of antenna-only is assumed to be zero. The price variation that is needed to estimate price impacts arises from the cable alternatives.

Table 1 provides information about the sampled households and the service options that are available to them. Nearly 85% of the sample lives in single-family dwellings, and average income is approximately $62,000. The most popular television option is basic and extended cable, which is chosen by 45% of the households. Less than a quarter of the households have antenna-only reception. The average price for basic and extended cable is approximately $28 per month, with this price ranging from $16 to $45 (not shown in the table). The fee for premium cable is $40 on average, ranging from $26 to $56.

Table 1
DEMOGRAPHIC VARIABLES AND SERVICE ATTRIBUTES

Average income	$62,368
Income Groups	*Share (%)*
<$25,000	19.60
$25,000–$49,999	24.48
$50,000–$74,999	24.39
$75,000–$99,999	17.44
>$100,000	14.09
Unmarried	31.93
Single-family dwelling	84.58
Rent	16.34
Household Size	
1 person	18.88
2 people	39.40
3 people	16.76
4 people	15.39
5 or more people	9.56
Chosen Television Option	
Antenna-only	23.14
Basic and extended cable	44.79
Premium cable	20.72
Satellite	11.36
Attributes of Service in Household's Area	*Average*
Over-the-air channels	10.7
Basic/extended cable channels	62.9
Additional premium cable channels	5.8
Price for basic and extended cable	27.96
Price for premium cable	39.58

More details of the data appear in the Web Appendix (http://www.marketingpower.com/jmrfeb10).

Because the attributes of the television alternatives are the same for all households in a geographic market, we add a subscript for markets. Let U_{njm} be the utility that household n that lives in market m obtains from alternative j. The price of alternative j in market m is p_{mj}, which is not subscripted by n because it is the same for all households in the market m. Price is zero for antenna-only television, and the price of satellite television does not vary over markets or households. The price of the two cable options varies over geographic markets, and unobserved attributes of cable service (e.g., quality of programming) are expected to be correlated with price. The utility of the two cable options $(j = 2, 3)$ is specified as in Example 2:

$$U_{njm} = V(p_{mj}, x_{mj}, \beta_n) + \varepsilon^1_{nj} + \varepsilon^2_{nj}, \tag{18}$$

where ε^1_{nj} is correlated with price, ε^2_{nj} is i.i.d. extreme value, and x_{nj} captures exogenous observed attributes. Utility for the two options with constant price $(j = 1, 4)$ is the same but without the correlated error component ε^1_{nj}. Price for the cable options is specified as linear in instruments plus a separable error:

$$p_{mj} = \gamma z_{mj} + \mu_{mj}. \tag{19}$$

We specify μ_{mj} and ε^1_{nj} for $j = 2, 3$ to be jointly normal, independent over j. Then, utility with the control function for alternative $j = 2, 3$ is as follows:

$$U_{njm} = V(p_{mj}, x_{mj}, \beta_n) + \lambda\mu_{mj} + \sigma_j\eta_{nj} + \varepsilon^2_{nj}, \tag{20}$$

where η_{nj} is standard normal.

To complete the model, we specify $V(\cdot)$ as follows:

$$V(p_{mj}, x_{mj}, \beta_n) = \alpha p_{mj} + \sum_{g=2\text{–}5} \theta_g p_{mj} d_{gn} + \tau x_{mj} + \delta_j k_j + \kappa k_j s_n + \varphi\omega_n c_j. \tag{21}$$

We specify the price effect to differ by income group. We identified five income groups and take the lowest-income group as the base. The dummy d_{gn} identifies whether household n is in income group g. The price coefficient for a household in the lowest-income group is α, while that for a household in group $g > 1$ is $\alpha + \theta_g$. The nonprice attributes x_{mj} enter with fixed coefficients. The alternative-specific constant for alternative j is k_j. These constants are entered directly and also interacted with demographic variables, s_n.

We include an error component to allow for correlation in unobserved utility over the three nonantenna alternatives. In particular, $c_j = 1$ if j is one of the three nonantenna alternatives, and $c_j = 0$ if otherwise, and ω_n is an i.i.d. standard normal deviate. The coefficient φ is the standard deviation of this error component, reflecting the degree of correlation among the nonantenna alternatives.

Therefore, the choice probability takes the form of a mixed logit (Brownstone and Train 1999; Train 1998), with the mixing over the distribution of the error components:

$$P_{ni} = \int \frac{e^{V_{ni}(\eta_2,\eta_3,\omega)}}{\sum_{j=1}^4 e^{V_{nj}(\eta_2,\eta_3,\omega)}} \varphi(\eta_2)\varphi(\eta_3)\varphi(\omega)\, d\omega d\eta_3 d\eta_2, \tag{22}$$

where $\varphi(\cdot)$ is the standard normal density and

$$(23)\quad V(\eta_j, \omega) = \alpha p_{mj} + \sum_{g=2-5} \theta_g p_{mj} d_{gn} + \tau x_{mj} + \delta_j k_j + \kappa k_j s_n + \varphi\omega c_j + \lambda \mu_{mj} + \sigma_j \eta_j.$$

The integral is approximated through simulation: A value of η_2, η_3, and ω is drawn from their standard normal densities, the logit formula is calculated for this draw, the process is repeated for numerous draws, and the results are averaged. To increase accuracy, we use Halton (1960) draws instead of independent random draws. Bhat (2001) finds that 100 Halton draws perform better than 1000 independent random draws, a result that has been confirmed on other data sets (see Hensher 2001; Munizaga and Alvarez-Daziano 2001; Train 2000, 2003).

RESULTS

The first step of the approach is to estimate the pricing functions to recover the residuals entering the control functions in the choice model. We regressed the price in each market against the product attributes listed in Table 2 plus Hausman-type price instruments (see Hausman 1997b). We calculated the price instrument for market m as the average price in other markets that are served by the same multiple-system operator as market m.[7] In our context, these instruments are appropriate if the prices of the same multiple-system operator in other markets reflect common costs of the multiple-system operator but not common demand shocks (e.g., unobserved advertising). We created a separate instrument for the price of extended basic cable and the price of premium cable, and we ran separate regressions for extended basic price and premium price using all instruments in each equation.[8]

The residuals from these regressions enter without transformation in the mixed logit model; that is, the control functions are a coefficient times the product-market residual, which is the first and simplest specification proposed in the "Model" section. Specifically, the residual from the extended basic cable price regression enters the extended basic cable alternatives and similarly for the premium cable.

Table 2 gives the estimated parameters. The variables are listed in three groups: (1) those that vary over markets but not over consumers in each market, (2) those that vary over consumers in each market, and (3) the extra variables that are included to correct for endogeneity. The first column gives the model without any correction for the correlation between price and omitted attributes; utility is the same as we specified previously except that the residuals, $\hat{\mu}_{mj}$, and induced error components, η_j, are not included. The second column applies the control function approach by including the residuals and error components.

Without correction, the base price coefficient α is estimated to be −.0202. As we stated previously, the price coefficient is allowed to differ by income group; the estimated price coefficient diffentials by income group, θ_g, are given in the second part of the table because these variables vary over households in each market. For the second income group, the estimated price coefficient is the base of −.0202 plus the differential of .0149, for an estimated price coefficient of −.0053. However, note that for Income Groups 3–5, which comprise the majority of households, the estimated differential exceeds the base in magnitude, such that the estimated price coefficients are positive. This result contradicts the expectation of downward-sloping demand and renders the model implausible for predictive purposes and welfare analysis (because welfare analysis assumes a negative price coefficient).

Inclusion of the control functions adjusts the estimated price coefficients in the expected way. We obtained a significant, negative price coefficient for all income groups because the base coefficient estimate increases almost fivefold to −.10. Price elasticities decrease as income rises, with the highest-income group obtaining a price coefficient that is approximately 30% smaller than that of the lowest-income group.

The residuals enter significantly and with the expected sign. In particular, a positive residual occurs when the price of the product is higher than can be explained by observed attributes and other observed factors. A positive residual suggests that the product possesses desirable attributes that are not included in the analysis. The residual entering the demand model with a positive coefficient is consistent with this interpretation.

Neither of the error components is statistically significant, and the hypothesis that both have zero standard deviations cannot be rejected at any meaningful level of confidence. This result might imply that the residuals capture the market-specific unobserved attributed of cable service completely or perhaps reflects the empirical difficulty of estimating alternative-specific normal error components when each alternative also has an i.i.d. extreme value error component.

Several product attributes are included in the model. In the model without correction, one of these attributes enters with an implausible sign: number of cable channels. With correction, all the product attributes enter with expected signs. In general, the magnitudes are reasonable. An extra premium channel is valued more than an extra cable (nonpremium) channel. An extra over-the-air channel is also valued more than an extra nonpremium cable channel, perhaps because the proliferation of cable channels with low programming content makes the value of extra cable channels relatively low. The option to obtain pay-per-view is valued highly. Note that this attribute, unlike the others, is not on a per-channel basis; its coefficient represents the value of the option to purchase pay-per-view events. The point estimates imply that households are willing to pay \$6.00–\$8.88 per month for this option, depending on their income.

Several demographic variables enter the model. Their estimated coefficients are fairly similar in the corrected and uncorrected models. The estimates suggest that households with higher education tend to purchase less television reception: The education coefficients are progressively more highly negative for antenna-only (which is zero by normalization), extended basic cable, premium cable, and

[7]To our knowledge, there has been no prior application of the control function approach with cross-sectional market data. Villas-Boas and Winer (1999) use lagged prices as instruments in their time-series model.

[8]We discuss the use of alternative instruments in the "Robustness Analysis" section.

Table 2
MIXED LOGIT MODEL OF TELEVISION RECEPTION CHOICE: CONTROL FUNCTION APPROACH

Explanatory Variable	*Uncorrected*	*With Control Function*
Variables That Vary over Markets but Are Constant over Consumers in Each Market		
Price, in dollars per month [1–4]	**−.0202** (.0047)	**−.1003** (.0471)
Number of cable channels [2, 3]	**−.0023** (.0011)	.0026 (.0039)
Number of premium channels [3]	**.0375** (.0163)	.0559 (.0382)
Number of over-the-air channels [1]	**.0265** (.0090)	.0232 (.0152)
Whether pay-per-view is offered [2, 3]	**.4315** (.0666)	**.5992** (.1792)
Indicator: AT&T is cable company [2]	.1279 (.0946)	−.2072 (.2437)
Indicator: AT&T is cable company [3]	.0993 (.1195)	−.2559 (.2737)
Indicator: Adelphia is cable company [2]	**.3304** (.1224)	.3443 (.2930)
Indicator: Adelphia is cable company [3]	.2817 (.1511)	.2504 (.3400)
Indicator: Cablevision is cable company [2]	**.6923** (.2243)	.1031 (.3749)
Indicator: Cablevision is cable company [3]	**1.328** (.2448)	1.015 (.5412)
Indicator: Charter is cable company [2]	.0279 (.1010)	−.0587 (.2259)
Indicator: Charter is cable company [3]	−.0618 (.1310)	−.2171 (.2139)
Indicator: Comcast is cable company [2]	**.2325** (.1107)	−.1111 (.3694)
Indicator: Comcast is cable company [3]	**.5010** (.1325)	.2619 (.3210)
Indicator: Cox is cable company [2]	**.2907** (.1386)	−.0720 (.3314)
Indicator: Cox is cable company [3]	**.5258** (.1637)	.1678 (.5065)
Indicator: Time Warner cable company [2]	.1393 (.0974)	−.0902 (.2213)
Indicator: Time Warner cable company [3]	.2294 (.1242)	−.0462 (.2254)
Alternative-specific constant [2]	**1.119** (.2668)	**3.060** (1.054)
Alternative-specific constant [3]	.1683 (.3158)	2.439 (1.542)
Alternative-specific constant [4]	−.2213 (.4102)	4.386 (2.690)
Variables That Vary over Consumers in Each Market		
Price for Income Group 2 [1–4]	**.0149** (.0024)	**.0154** (.0026)
Price for Income Group 3 [1–4]	**.0246** (.0030)	**.0253** (.0038)
Price for Income Group 4 [1–4]	**.0269** (.0034)	**.0271** (.0042)
Price for Income Group 5 [1–4]	**.0308** (.0036)	**.0311** (.0040)
Education level of household [2]	**−.0644** (.0220)	**−.0640** (.0254)
Education level of household [3]	**−.1137** (.0278)	**−.1129** (.0371)
Education level of household [4]	**−.1965** (.0369)	**−.1987** (.0384)
Household size [2]	**−.0494** (.0240)	**−.0556** (.0283)
Household size [3]	.0160 (.0286)	.0303 (.0421)
Household size [4]	.0044 (.0357)	.0023 (.0447)
Household rents dwelling [2–3]	**−.2471** (.0867)	**−.2719** (.0891)
Household rents dwelling [4]	−.2129 (.1562)	−.2008 (.1329)
Single-family dwelling [4]	**.7622** (.1523)	**.7790** (.2071)
Error component for nonantenna alternatives, SD [2–4]	.5087 (.6789)	.4994 (.7344)
Terms to Correct for Endogeneity		
Residual for extended basic cable price [2]		.0833 (.0481)
Residual for premium cable price [3]		**.0929** (.0499)
Error component for basic and extended cable, SD [2]		.0488 (1.423)
Error component for premium cable, SD [3]		1.425 (1.142)
Log-likelihood at convergence	−14,660.84	−14,645.21
Number of observations	11,810	11,810

Notes: Alternatives: (1) antenna-only, (2) basic and extended cable, (3) premium cable, and (4) satellite. Each variable enters the alternatives that are listed in brackets and takes the value of zero in other alternatives. Estimates that are statistically significant ($p < .05$) are in bold. Standard errors are in parentheses.

satellite. Larger households tend not to buy extended basic cable as readily as smaller households. Differences by household size with respect to the other alternatives are highly insignificant. We included a dummy for whether the household rents its dwelling in the two cable alternatives and separately in the satellite alternative. These variables account for the notion that renters are perhaps less able to install a cable hookup and less willing to incur the capital cost of a satellite dish than a household that owns its dwelling. The estimated coefficients are negative, confirming these expectations. Finally, a dummy for whether the household lives in a single-family dwelling enters the satellite alternative to account for the relative difficulty in installing a satellite dish on a multifamily dwelling. As we expected, the estimated coefficient is positive.

Fewer coefficients are significant in the model with correction for endogeneity than in the uncorrected one. We expected this result because the correction for endogeneity attempts to obtain more information from the data (i.e., the relationship of unobserved factors to price and the relationship of observed factors to demand). In other words, the uncorrected model gives a false sense of precision by assuming that price is independent of unobserved factors, when indeed price is related to these factors. Notably, all the coefficients that become insignificant with correction, when they were significant without correction, are for variables that vary over markets but not over consumers in each market. This pattern reflects that unobserved attributes that are correlated with price vary over markets but not over consumers within each market because price itself only varies over markets.

ROBUSTNESS ANALYSIS

The appropriate control function and distribution for $\tilde{\varepsilon}_{jn}$ is a specification issue. We tried other specifications, including both residuals entering in each cable alternative (to allow for correlation across alternatives as in example 3); a series expansion, both signed and unsigned, of the residuals (to allow for the conditional mean not being exactly as given by a joint normal); correlated rather than independent error components; and exclusion of one or both of the error components (because they are not significant). These alternative specifications all provided similar results.

As is always the case with endogeneity, the selection of instruments is an issue. As we stated previously, we used the product attributes and Hausman-type prices as instruments, which are widely used but controversial (Bresnahan 1997; Hausman 1997a). With disaggregate demand models, the need for additional instruments is not as stringent as in models with just aggregate data because aggregate demographics do not enter the disaggregate models, but they affect market price. Therefore, they can serve as the extra instruments that are needed for demand estimation.[9]

We reestimated the model without using the prices in other areas as instruments but including the aggregate demographics. With the control function approach, the estimated price coefficient rose when we removed the Hausman-type prices as instruments. This is the direction of change that would be expected if the prices in other markets incorporated the impact of unobserved demand shocks. The other coefficients were not affected under either approach.

COMPARISON WITH PRODUCT-MARKET CONTROL

Given the widespread use of the BLP approach, we provide a brief comparison with the control function approach. Then, we discuss BLP model estimation and the results for the same data.

The BLP approach uses the aggregate demand equations to recover the unobserved demand factors by matching observed market shares to those predicted by the model. In contrast, our control function approach is based on using different equations for information on the unobserved demand factor, such as the pricing or advertising equation.

In most applications, the control function approach will be easier to implement than the BLP approach. Often, the first step is just a regression, and the second is maximum likelihood, so the approach can be estimated with standard software packages, such as STATA, SAS (which now has a mixed logit and probit routine), LIMDEP, and Biogeme.[10] The two-step estimator requires us to account for the estimated regressors (as discussed previously), and the sampling covariance can be estimated by bootstrap with these packages.

It is necessary to incorporate a contraction procedure into the estimation routine to implement the BLP estimator. It iteratively calculates the constants that equate predicted and actual shares at each trial value of the parameters. This computation is not trivial, especially when consumer-level data are being used in the estimated specification. Because of this computational burden, the BLP procedure is, to our knowledge, still not available in any of the common statistical packages.

Because the BLP approach matches observed to predicted shares in a nonlinear setting, it turns out to be sensitive to sampling error in market shares, as Berry, Linton, and Pakes (2004) show. It is not consistent in settings in which there are zero, one, or just a small number of purchase observations per product relative to the number of consumers, as is the case in some data sets.[11] It also requires all goods to be strict substitutes, something not required by the control function setup.

The BLP approach includes a constant δ_{mj} for each alternative in each market. All the elements of utility that do not vary within a market are subsumed into these constants. The utility specification given previously becomes the following:

$$U_{njm} = \delta_{mj} + \sum_{g=2-5} \theta_g p_{mj} d_{gn} + \kappa k_j s_n + \varphi \omega_n c_j + \varepsilon^2_{nj}. \tag{24}$$

The constants are expressed as a function of price and other observed attributes:

$$\delta_{mj} = \alpha p_{mj} + \tau x_{mj} + \delta_j k_j + \varepsilon^1_{mj}. \tag{25}$$

Assuming ε^2_{nj} and ω_n are i.i.d. extreme value and standard normal, respectively, leads to a mixed logit of the same form as for the control function approach except with constants for each product-market alternative and without the extra error components that are induced by the control function. We estimate the equation for the constants by instrumental variables because utility is assumed to be linear in ε^1_{mj} and ε^1_{mj} is correlated with price.

We perform estimation in two stages, with the first stage being the computationally burdensome one. First, we estimate the mixed logit model with constants for each alternative and each market. We recover these constants by

[9]Consider two households that have the same demographics but live in areas in which the aggregate demographics are different. Part of the price difference between the two areas is presumably attributable to the difference in aggregate demographics. This part of the price difference provides variation in price over households that can be used for estimation of price response.

[10]Matlab and Gauss codes for mixed logit are also available (free) from Train's Web site at http://elsa.berkeley.edu/~train/software.html.

[11]For example, in many housing data sets, houses are purchased zero or one time, violating the consistency condition of the BLP estimator. Another example is Martin's (2008) study of customers' choice between incandescent and compact fluorescent light bulbs (CFLs), where advertising and promotions (e.g., discount coupons) occurred on a weekly basis and varied over stores, and yet it was common for a store not to sell any CFLs in a given week. With the market defined as a store–week, Martin reports that 65% of the market shares were zero for CFLs.

Table 3
MIXED LOGIT MODEL OF TV RECEPTION CHOICE: BLP APPROACH

Explanatory Variable	*Ordinary Least Squares*	*Three-Stage Least Squares*
Variables That Vary over Markets but Are Constant over Consumers in Each Market		
Price, in dollars per month [1–4]	−**.0245** (.0091)	−**.0922** (.0409)
Number of cable channels [2, 3]	−.0024 (.0027)	.0017 (.0042)
Number of premium channels [3]	.0132 (.0502)	.0463 (.0329)
Number of over-the-air channels (negative) [1]	.0168 (.0132)	.0196 (.0186)
Whether pay-per-view is offered [2, 3]	**.5872** (.1326)	**.7144** (.1814)
Indicator: AT&T is cable company [2]	−.3458 (.2127)	−.2934 (.2353)
Indicator: AT&T is cable company [3]	.0158 (.2262)	−.0017 (.2541)
Indicator: Adelphia Comm is cable company [2]	.4883 (.2943)	.3837 (.2733)
Indicator: Adelphia Comm is cable company [3]	.6111 (.3121)	.5219 (.3065)
Indicator: Cablevision is cable company [2]	.1905 (.5368)	−.1912 (.5596)
Indicator: Cablevision is cable company [3]	**1.215** (.5829)	.7400 (.6193)
Indicator: Charter Comm is cable company [2]	−.1807 (.2387)	−.1871 (.2196)
Indicator: Charter Comm is cable company [3]	−.0408 (.2539)	−.0685 (.2488)
Indicator: Comcast is cable company [2]	−.4097 (.2601)	−.4034 (.2755)
Indicator: Comcast is cable company [3]	.1427 (.2755)	.0989 (.3002)
Indicator: Cox Comm is cable company [2]	−.6419 (.4302)	−.6336 (.4225)
Indicator: Cox Comm is cable company [3]	−.0398 (.4564)	−.1563 (.4827)
Indicator: Time Warner is cable company [2]	−.3756 (.2335)	−.3439 (.2281)
Indicator: Time Warner cable company [3]	.0527 (.2503)	−.0009 (.2597)
Alternative-specific constant [2]	**1.659** (.3486)	**3.185** (1.007)
Alternative-specific constant [3]	.6462 (.4725)	2.819 (1.480)
Alternative-specific constant [4]	**.6583** (.1733)	**4.635** (.2193)
Variables That Vary over Consumers in Each Market		
Price for Income Group 2 [1–4]	**.0156** (.0021)	
Price for Income Group 3 [1–4]	**.0273** (.0023)	
Price for Income Group 4 [1–4]	**.0299** (.0027)	
Price for Income Group 5 [1–4]	**.0353** (.0029)	
Education level of household [2]	−**.0521** (.0173)	
Education level of household [3]	−**.1385** (.0203)	
Education level of household [4]	−**.2525** (.0308)	
Household size [2]	−**.0984** (.0240)	
Household size [3]	−.0155 (.0277)	
Household size [4]	−.0235 (.0363)	
Household rents dwelling [2, 3]	−**.1494** (.0772)	
Household rents dwelling [4]	−.5470 (.1349)	
Single-family dwelling [4]	.1967 (.1023)	
Error component for nonantenna alternatives, SD [2–4]	**.7775** (.1664)	
Log-likelihood at convergence	−13,927.40	
Number of observations	11,810	

Notes: Alternatives: (1) antenna-only, (2) basic and extended cable, (3) premium cable, and (4) satellite. Each variable enters the alternatives that are listed in brackets and takes the value zero in other alternatives. Estimates that are statistically significant ($p < .05$) are in bold. Standard errors are in parentheses.

solving for the values that match observed to predicted market shares in each market and for each product at every set of parameter values until the minimum is located. Then, these estimated constants are regressed against the product attributes using three-stage least squares.[12] A separate equation is used for the extended basic cable, premium cable, and satellite constants, with the coefficients of the product attributes constrained across equations as in the control function setup (and every characteristics-based setup of which we are aware).

The results appear in Table 3. The bottom part of the table gives the estimates of the demographic coefficients in the mixed logit model. The top part of the table gives the results of the regression of constants on product attributes. The first column at the top gives the ordinary least squares results, which do not account for omitted attributes, and the second column gives the three-stage least squares results.

As with the control function approach, the correction for omitted variables raises the estimated price coefficients. Without correction, three of the five income groups received a positive estimated price coefficient. With correction, all groups obtain a significantly negative price coefficient.

The estimated base price coefficient is −.0922, compared with the −.1003 obtained with the control function

[12] The negative of the number of over-the-air channels enters these equations because this attribute enters the antenna-only alternative in the model of Table 2, whereas it is now entering the constants for the nonantenna alternatives.

Table 4
ESTIMATED ELASTICITIES

	Control Function	*BLP*
Price of Extended Basic Cable		
Antenna-only share	.97	.79
Extended basic cable share	−1.08	−.97
Premium cable share	.76	.88
Satellite share	1.02	.87
Price of Premium Cable		
Antenna-only share	.48	.52
Extended basic cable share	.50	.57
Premium cable share	−1.83	−2.04
Satellite share	.48	.58
Price of Satellite		
Antenna-only share	.50	.42
Extended basic cable share	.40	.43
Premium cable share	.37	.45
Satellite share	−3.77	−3.59

approach. The estimates of θ_g, the incremental price coefficient for higher-income groups, are similar under the two approaches. As in the control function approach, the number of cable channels obtains a negative coefficient when endogeneity is ignored and, as we expected, becomes positive when the endogeneity is corrected. All the product attributes obtain similar values as with the control function approach.

The demographic coefficients in Table 3 provide similar conclusions as those from the control function approach. Education induces households to buy less television reception. Larger households tend not to buy extended basic cable. Renters tend not to buy cable and satellite as readily as owners. Single-family dwellers tend to purchase satellite reception more readily than households that live in multifamily dwellings. Differences appear not to be statistically significant.

Table 4 gives price elasticities from the models for each approach. The two methods give similar elasticities. For example, the same-price elasticity for basic and extended cable is −1.08 with the control function approach and −.97 under the BLP approach.

CONCLUSION

The concern that price, advertising, or other variables are endogenous has proved to be important in many applications. In this article, we propose a control function approach for handling endogeneity in choice models. It uses observed variables and economic theory to derive controls for the part of the unobserved demand factor that is not independent of the endogenous variable. We use the pricing equation in this article to derive controls, but we believe that there are many other possible equations (e.g., the advertising equation) that also contain information on unobserved demand factors.

The approach provides an alternative to the commonly used BLP product-market controls for unobserved quality, which is sensitive to sampling error in market shares, more difficult to estimate (especially with consumer-level data), and not applicable for many recently proposed discrete choice demand estimators.

We apply both methods to examine households' choices among television options, including basic and premium cable packages, in which unobserved attributes, such as quality of programming, are expected to be correlated with price. Without correcting for endogeneity, aggregate demand for each television option is upward-sloping. The corrected estimates from both the control function method and the product-market controls method produce similar and much more realistic demand elasticities.

REFERENCES

Bajari, Patrick, Jeremy Fox, Kyoo Il Kim, and Stephen Ryan (2007), "A Simple Nonparametric Estimator for the Distribution of Random Coefficients in Discrete Choice Models," working paper, Department of Economics, University of Minnesota.

Bass, Frank (1969), "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research*, 6 (August), 291–300.

Berry, Steve (1994), "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25 (2), 242–62.

———, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–89.

———, ———, and ——— (2004), "Differentiated Products Demand System from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112 (1), 68–105.

———, Oliver Linton, and Ariel Pakes (2004), "Limits Theorems for Estimating Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, 71 (3), 613–54.

Bhat, Chandra (2001), "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research B*, 35 (7), 677–93.

Blundell, Richard and James Powell (2004), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71 (3), 655–79.

Bresnahan, Tim (1997), "The Apple-Cinnamon Cheerios War: Valuing New Goods, Identifying Market Power, and Economic Measurement," working paper, Department of Economics, Stanford University.

Brownstone, David and Kenneth Train (1999), "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, 89 (1–2), 109–129.

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

Crawford, Gregory (2000), "The Impact of the 1992 Cable Act on Household Demand and Welfare," *RAND Journal of Economics*, 31 (3), 422–49.

Fox, Jeremy (2008), "Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices," *RAND Journal of Economics*, 38 (4), 1002–1019.

Goolsbee, Austan and Amil Petrin (2004), "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica*, 72 (2), 351–82.

Gupta, Sachin and Sungho Park (2009), "A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data," *Journal of Marketing Research*, 46 (August), 531–42.

Halton, John (1960), "On the Efficiency of Evaluating Certain Quasi-Random Sequences of Points in Evaluating Multi-dimensional Integrals," *Numerische Mathematik*, 2 (1), 84–90.

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

——— (1997a), "Reply to Professor Bresnahan," working paper, Department of Economics, Massachusetts Institute of Technology.

——— (1997b), "Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, R. Gordon and T. Bresnahan, eds. Chicago: University of Chicago Press, 207–248.

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

——— and Richard Robb (1985), "Alternative Methods for Evaluating the Impacts of Interventions: An Overview," *Journal of Econometrics*, 30 (1–2), 239–67.

Hendel, Igal and Aviv Nevo (2006), "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74 (6), 1636–74.

Hensher, David (2001), "The Valuation of Commuter Travel Time Savings for Car Drivers in New Zealand: Evaluating Alternative Model Specifications," *Transportation*, 28 (2), 101–118.

Imbens, Guido and Whitney Newey (2008), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," working paper, Department of Economics, Harvard University.

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

Karaca-Mandic, Pinar and Kenneth Train (2003), "Standard Error Correction in Two-Step Estimation with Nested Samples," *Econometrics Journal*, 6 (2), 401–407.

Kuksov, Dmitri and J. Miguel Villas-Boas (2008), "Endogeneity and Individual Customer Choice," *Journal of Marketing Research*, 45 (December), 702–714.

Louviere, Jordan, Kenneth Train, Moshe Ben-Akiva, Chandra Bhat, David Brownstone, Trudy Cameron, et al. (2005), "Recent Progress on Endogeneity in Choice Modeling," *Marketing Letters*, 16 (3–4), 255–65.

Martin, Leslie (2008), "Consumer Demand for Compact Fluorescent Light Bulbs," working paper, Department of Agricultural and Resource Economics, University of California, Berkeley.

Munizaga, Marcela and Ricardo Alvarez-Daziano (2001), "Mixed Logit Versus Nested Logit and Probit," working paper, Departmento de Ingeniera Civil, Universidad de Chile.

Murphy, Kevin and Robert Topel (1985), "Estimation and Inference in Two Step Econometric Models," *Journal of Business and Economic Statistics*, 3 (4), 370–79.

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

Newey, Whitney and Daniel McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in *The Handbook of Econometrics*, Vol. 4, Z. Griliches and M. Intriligator, eds. Amsterdam: Elsevier Science Publishers, 2111–45.

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

Rivers, Douglas and Quang Vuong (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39 (3), 347–66.

Smith, Richard and Richard Blundell (1986), "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54 (3), 679–86.

Train, Kenneth (1998), "Recreation Demand Models with Taste Differences over People," *Land Economics*, 74 (2), 230–39.

——— (2000), "Halton Sequences for Mixed Logit," Working Paper No. E00-278, Department of Economics, University of California, Berkeley.

——— (2003), *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.

Villas-Boas, J. Miguel (2007), "A Note on Limited Versus Full Information Estimation in Non-Linear Models," working paper, Haas School of Business, University of California, Berkeley.

——— and Russell Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45 (10), 1324–38.

Yang, Sha, Yuxin Chen, and Greg Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1 (3), 251–75.