Omitted Product Attributes in Discrete Choice Models

by
Amil Petrin and Kenneth Train
University of Chicago and University of California, Berkeley

August 9, 2002

Abstract

We describe two methods for correcting for omitted variables in discrete choice models: a fixed effects approach and a control function approach. We apply the methods to disaggregate data on customer's choice among television options including cable, satellite, and antenna. Estimates are similar for the two methods, and the estimated price response rises substantially when the correction is applied with either method.

1 Introduction

Models of differentiated products are widely used for merger analysis (where elasticities and cross-elasticities among similar products determine the welfare implications of a merger), marketing (where the demand for one product depends on the attributes of all similar products), policy analysis (where impacts often depend on substitution patterns, such as whether the induced demand for new energy-efficient vehicles is drawn more from "gas-guzzlers" or current "gas-sippers"), design and forecasting of new products (where demand depends on the new product's similarity to other products, and where the issue of self-cannibalization of the firm's similar products is critical for profits), and a host of other issues.

In aggregate (i.e., market-level) models of differentiated products, price is usually endogeneous, determined by the interaction of demand and supply. Since the demand for differentiated products under heterogeneous preferences is inherently non-linear, the application of standard methods for correcting for this endogeneity are not immediately applicable. Berry, Levinsohn, and Pakes (1995, henceforth BLP) developed and applied a method, utilizing concepts of Berry (1994), that provides consistent estimation in the face of market endogeneity. The method has proven successful in numerous applications, including the demand for cable TV (Crawford, 2000), cereals (Nevo, 2001), and minivans (Crawford, 2002), to name only a few.

With disaggregate (i.e., customer-level) models of demand, price is not necessarily endogenous in the traditional sense, since the demand of the customer does not usually affect market price. However, the same issues that give rise to the need for correction in aggregate models can often appear in disaggregate models. In particular, omitted product attributes can create correlation between the price and the unobserved portion of utility: the market mechanism causes the price to be higher for products that display desirable attributes that are observed by consumers but not measured by the econometrician. Since these attributes affect demand at the customer level, price is correlated with the error term even in disaggregate demand models. The BLP approach can be applied to disaggregate data, or a combination of aggregate and disaggregate data, as illustrated by Berry et al. (1998) and Goolsbee and Petrin (2002).

An alternative approach, based on control functions, is described by Blundell and Powell (2001) for general non-linear models, expanding concepts that date back to Heckman (1978) and Hausman (1978). The term was introduced by Heckman and Robb (1985) in the context of selection models. The method has been applied to a Tobit model by Smith and Blundell (1986) and probit by Villas-Boas and Winer (1999). The control function approach can sometimes be easier to implement than BLP's approach and can be used in situations for which the BLP approach is infeasible. However, it involves more stringent assumptions. For consistency, the instruments in the control function approach must be independent of all the remaining errors terms, while the instruments in BLP's method need only be *mean* independent of just the errors that enter the fixed effects regressions. When the stronger assumption holds, the control function approach is more efficient. In the sections below we describe both approaches and apply them to disaggregate data on customers' choice among TV options. We find that the two approaches provide very similar estimates in this application, including a substantial increase in the price response when either correction is applied.

2 Specification

We use notation that adapts well for our application. There are J products indexed by j. The price and attributes of the products vary over M markets. The price and some product attributes are observed by the econometrician; these are denoted p_{mj} and x_{mj}, respectively, for product j in market m. Some attributes are not observed by the econometrician but are known by consumers and affect their demand. Customer n buys in one of the markets; for simplicity, we say that customer n buys in market m without explicitly denoting the fact that m differs for different n. The more correct but more cumbersome notation would be $m(n)$ as the market in which n buys. The utility that customer n who lives in market m obtains from product j is decomposed into observed and unobserved parts:

$$U_{nj} = V(p_{mj}, x_{mj}, s_n) + e_{nj} \tag{1}$$

where s_n denotes the observed characteristics of the customer, V is a calculable function up to parameters, and e_{nj} is defined as the difference that makes the equation an identity.

The choice probability is defined in the traditional way. Let e_n denote the vector $\langle e_{n1}, \ldots, e_{nJ} \rangle$, and let $\varphi(de_n)$ denote the density of e_n conditional on the observed variables. The choice probability for good i is then

$$P_{ni} = \int_{A_{ni}} \varphi(de_n) \tag{2}$$

where $A_{ni} = \{e_n \mid U_{ni} > U_{nj} \; \forall j \neq i\}$ is the set of e_n such that product i provides maximal utility.

The difficulty arises because $\varphi(\cdot)$ does not take a convenenient form due to the correlation of e_{nj} with p_{mj}. The standard choice models are derived under assumptions that do not incorporate this correlation. Simple logit (e.g., McFadden, 1974) assumes that the unobserved component of utility is independent of the observed variables. Mixed logit and probit (e.g., Brownstone and Train, 1999) allow the distribution of the unobserved component to depend on observed variables; however, this dependence takes a particular form, such as arises with random coefficients, that does not reflect the type of correlation induced by omitted variables.

2.1 Control function approach

The basic concept motivating the control function approach is that the part of price that cannot be explained by observed attributes contains information about the value of the unobserved attributes. Price in each market is expressed in reduced form as a function of observed variables:

$$p_{mj} = g(j, z_m) + \mu_{mj},$$

where instruments z_m include the observed attributes of the products and other observed variables that are independent of μ_{mj}. The error μ_{mj} incorporates factors that affect price but are not captured by z_m, including the average value of the unobserved attributes of the products.

The unobserved component of utility in (1) is decomposed:

$$e_{nj} = f_j(\mu_m) + \varepsilon_{nj}, \tag{3}$$

where $f_j(\mu_m)$ captures the mean of e_{nj} conditional on $\mu_m = \langle \mu_{m1}, \ldots, \mu_{mJ} \rangle$. f_j is called the control function for alternative j, and the elements of μ_m are called the control variables. Given p_{mj} and z_m $\forall j, m$, the values of the control functions are observed and can be included in the observed portion of utility. Substituting (3) into (1) gives:

$$U_{nj} = V(p_{mj}, x_{mj}, s_n) + f_j(\mu_m) + \varepsilon_{nj}. \tag{4}$$

The choice probability is defined conditional on the control functions. Let ε_n denote the vector $\langle \varepsilon_{n1}, \ldots, \varepsilon_{nJ} \rangle$, and let $\phi(d\varepsilon_n)$ denote the density of ε_n conditional on the original observed variables and the control functions. The choice probability is then

$$P_{ni} = \int_{B_{ni}} \phi(d\varepsilon_n)$$

where $B_{ni} = \{\varepsilon_n \mid U_{ni} > U_{nj} \ \forall j \neq i\}$. This expression is conditional on μ_m while the probability in (2) is marginal over μ_m.

The density of ε_n can have a convenient form even if the density of e_n does not. For example, if ε_{nj} is independent of the explanatory variables entering V and any additional instruments entering z_{mj}, then P_{ni} takes the form of a standard logit or probit depending on the distribution. Note that the instruments need to be independent of both μ_n and ε_n. If ε_{nj} depends on the explanatory variables in a way that arises from random coefficients, then P_{ni} becomes a mixed logit or probit. Estimation proceeds in two stages. First, the prices are regressed against instruments, and the residuals are calculated. Then, the discrete choice model is estimated in the usual way with functions of these residuals entering as explanatory variables.

The purpose of the control function is to absorb the part of unobserved utility that is correlated with prices; once this part of unobserved utility is included, the remaining part of unobserved utility is not correlated with prices. Ideally, the control functions serve as proxies for the omitted attributes. Usually, as in Smith/Blundell and Villas-Boas/Winer, the control functions are linear in the residuals μ_m. However, any function can be specified: the issue is simply whether, given $f_j(\mu_m)$, the remaining unobserved portion of utility has a convenient density. Blundell and Powell discuss semi- and non-parameteric ways of estimating $f_j(\mu_m)$. Even though our notation does not indicate this possibility explicitly, the control functions can themselves have random coefficients, to reflect variation in how customers respond to the omitted attributes.

The control function approach can be interpreted in two ways. The first interpretation, as described above, is that the control functions proxy for the omitted attributes. A second interpretation arises from analogy to two-stage least squares, as follows. In a linear model, the control function approach is the same (that is, gives exactly the same estimates) as 2SLS. The model is

$$y_n = \alpha p_n + \beta x_n + e_n$$

where p_n is correlated with e_n. Under both approaches, the first step is to estimate $p_n = \gamma z_n + \mu_n$ where z_n are exogenous instruments including x_n. For 2SLS, the original equation is then estimated with the predicted price, $\hat{p}_n$, used in lieu of the actual price. Under the control function approach, the residual $\hat{\mu}_n$ is added as an explanatory variable in the original equation while retaining the actual price:

$$y_n = \alpha p_n + \beta x_n + \lambda \hat{\mu}_n + \varepsilon_n.$$

Substituting $\hat{p}_n + \hat{\mu}_n$ for p_n, the equation becomes

$$y_n = \alpha \hat{p}_n + \beta x_n + (\lambda + \alpha)\hat{\mu}_n + \varepsilon_n.$$

This last equation is the same as the second stage of 2SLS except that the first-stage residual enters as an extra explanatory variable. As Hausman (1978) notes, this residual is, by construction, orthogonal in the sample to $\hat{p}_n$ and x_n. Its inclusion therefore has no effect on the estimates of α and β.

Note that the control function approach can be applied, in principle, whenever observed variables entering utility are correlated with unobserved utility. For example, price might vary over all observations rather than over markets (i.e., groups of observations). As evident below, the fixed effects approach is feasible only when disaggregate observations are groupable and the endogeneity arises over groups, like markets, such that group-specific fixed effects are identifiable.

2.2 Fixed effects approach

The BLP approach is equivalent to including a separate set of alternative-specific constants for each market. Utility is decomposed into a part that is the same for all customers within a market, labeled δ_{mj}, plus observed and unobserved parts that vary over customers within the market:

$$U_{nj} = \delta_{mj} + \tilde{V}(p_{mj}, x_{mj}, s_n) + \tilde{\varepsilon}_{nj}.$$

Importantly, δ_{mj}, which is called the fixed effect for product j in market m, incorporates the average value of the omitted attributes along with the other components of utility that do not vary within a market. The unobserved portion of utility, $\tilde{\varepsilon}_{nj}$, is conditional on the average value of the omitted variables and, as such, need not be correlated with price.

The choice probability is

$$P_{ni} = \int_{C_{ni}} \psi(d\tilde{\varepsilon}_n)$$

where $C_{ni} = \{\tilde{\varepsilon}_n \mid U_{ni} > U_{nj} \ \forall j \neq i\}$ and $\psi(\cdot)$ is the density of $\tilde{\varepsilon}_n = \langle \tilde{\varepsilon}_{n1}, \ldots, \tilde{\varepsilon}_{nJ} \rangle$. This density can take any of the standard forms, since the correlation that gave rise to correlation with price is removed through the inclusion of δ_{mj}.

The fixed effect for each product in each market depends on its attributes in that market. This relation is assumed to be separable in price, observed attributes, and omitted attributes:

$$\delta_{mj} = \alpha p_{mj} + h(x_{mj}) + \xi_{mj}$$

As in the original utility specification, ξ_{mj} is correlated with p_{mj} since omitted attributes affect market prices. However, unlike the discrete choice model, the equation for δ_{mj} is linear. Standard 2SLS and 3SLS estimation can be readily applied. Consistency is attained if the instruments are mean independent of ξ_{mj}.

Estimation is performed in two steps. First, the discrete choice model is estimated. This model includes a constant for each product (except one, for normalization) in each market, for a total of $M \cdot (J-1)$ fixed effects. The model also includes interactions of demographic variables with observed attributes, which constitute elements of $\tilde{V}(\cdot)$. Given the large number of parameters, estimation of this model can be difficult computationally if the usual optimization methods are employed; however, procedures can be used, described below, that make the estimation relatively straightforward. Second, the fixed effects are regressed against product attributes. 3SLS is performed to account for the correlation of omitted attributes with price and the covariance among the fixed effects.

As stated, the key to implementing this method is estimation of the discrete choice model with fixed effects for each product and market. A computational device can be used to facilitate estimation. The parameters that enter $\tilde{V}$ are estimated by the standard optimization methods. At each trial value of these parameters, fixed effects are calculated that induce the forecasted shares in each market to equal the sample shares in that market. Berry (1994) shows that such fixed effects exist. The fixed effects are calculated iteratively by repeated application of the formula:

$$\delta_{mj}^{t+1} = \delta_{mj}^{t} + ln(S_{mj}) - ln(F_{mj}^{t}).$$

where t denotes the iteration, S_{mj} is the sample share for product j in market m, and F_{mj}^{t} is the forecasted share for product j in market m calculated with δ_{mj}^{t} $\forall j$. Note that this iteration for the fixed effects is performed for *each* iteration of the parameters that enter $\tilde{V}$. This procedure is called a contraction (BLP) or calibration (Train, 1986.)[1]

3 Application

We apply the methods to households' choice of television reception options. The specification and data are similar to those of Goolsbee and Petrin (2002). Four alternatives are considered available to households: (1) antenna only, (2) cable with basic or extended service, (3) cable with a premium service added, such as HBO, and (4) satellite dish. Basic and extended cable are combined because the data do not differentiate which of these options the households chose. Goolsbee and Petrin describe the market for cable and satellite TV, emphasizing the importance of accounting for omitted attributes, such as the quality of programming, in demand estimation. They applied the fixed effects approach, using data from 1999. We use data from 2001 that contain somewhat more information on households and the attributes of the alternatives. We apply both the fixed effects and the control function approach.

[1]When utility contains an additive extreme value error, this methodology is guaranteed to converge; otherwise, it is not. Other approaches can be used instead; see Goolsbee and Petrin for an approach that is effective when the error is multivariate normal.

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 a number of multiple system operators like 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. Details of the data are given in the appendix.

For the control function approach, utility is specified as:

$$U_{nj} = \alpha p_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} d_{gn} + \beta x_{mj} + k_j s_n + \lambda_j r_{mj} + \sigma \nu_n c_j + \epsilon_{nj}. \quad (5)$$

The price effect is specified to differ by income group. Five income groups are identified, with the lowest income group taken 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 alternative-specific constant for alternative j is k_j. These constants are interacted with demographic variables as well as entering directly. The variable r_{mj} is the residual from the first-stage price regression, for j representing either extended-basic cable and premium cable. No such residuals are included for antenna-only and satellite since these prices do not vary geographically. These residuals are the control functions that account for omitted attributes; we discuss their construction and alternative specifications below.

An error component is included to allow for correlation in unobserved utility over the three non-antenna alternatives. In particular, $c_j = 1$ if j is one of the three non-antenna alternatives and $c_j = 0$ otherwise, and ν_n is an iid standard normal deviate. The coefficient σ is the standard deviation of the error component, reflecting the degree of correlation among the non-antenna alternatives.

The final error term, ϵ_{nj}, is assumed to be iid extreme value, conditional on the explanatory variables including the control functions.[2] The choice probability therefore takes the form of a mixed logit (Train, 1998; Brownstone and Train, 1999), with the mixing over the distribution of ν_n:

$$P_{ni} = \int \frac{e^{V_{ni} + \sigma \nu c_i}}{\sum_{j=1}^{4} e^{V_{nj} + \sigma \nu c_j}} h(\nu) d\nu$$

[2]Note that ε_{nj} in (4) is the sum of the two error terms, $\sigma \nu_n c_j + \epsilon_{nj}$, in (5).

where $h(\cdot)$ is the standard normal density and $V_{nj} = \alpha p_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} d_{gn} + \beta x_{mj} + k_j s_n + \lambda_j r_{mj}$. The integral is approximated through simulation: a value of ν is drawn from the standard normal density, the logit formula is calculated for this value of ν, the process is repeated for numerous draws, and the results are averaged. To increase accuracy, Halton (1960) draws are used instead of independent random draws. Bhat(2001) found that 100 Halton draws perform better than 1000 independent random draws, a result that has been confirmed on other datasets by Train (2000, 2002), Hensher (2001), and Munizaga and Alvarez-Daziano (2001).

Table 1 gives the estimated parameters. The first column gives the model without any correction for the correlation between price and omitted attributes; utility is the same as specified above except that the residuals, r_{mj}, are not included. The second column applies the control function approach by including the residuals.[3] Without correction, the base price coefficient α is small, sufficiently so that the price coefficient $\alpha + \theta_g$ is positive for three of the five income groups, rendering the model implausible and unuseable for policy analysis. Inclusion of the control functions raises the magnitude of the estimated base price coefficient, as expected. A negative price coefficient is obtained for all incomes groups. The magnitude decreases as income rises, with the highest income group obtaining a price coefficient that is about thirty percent smaller than that of the lowest income group.

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 of the product attributes enter with expected signs. The magnitudes are generally reasonable. An extra premium channel is valued more than an extra cable (non-premium) channel. An extra over-the-air channel is also valued more than an extra non-premium cable channel, presumably because there are fewer over-the-air channels such that each one becomes more valuable. Stated differently,

[3]Since the residuals are estimates instead of true, the standard errors that are produced by the traditional formulas, and are output from the mixed logit estimation routines, are biased downward. We calculated standard errors by adding a bootstrap on the price regressions. That is, we repeatedly estimated the price regressions with bootstrapped samples, calculated the residuals, and estimated the mixed logit model with the new residuals. The variance in the logit estimates over the bootstrapped price samples was added to the sampling variance that is calculated for the logit estimates under fixed explanatory variables. These total standard errors are given in the table. This adjustment is important, especially for the base price coefficient and the coefficients for the residuals, whose standard errors doubled when the estimation variance in the residuals was reflected. The standard errors of the other product attribute coefficients rose by around fifty percent, and there was negligible effect for demographic coefficients.

Table 1: Mixed Logit Model of TV Reception Choice
Control Function Approach

Explanatory variable	Uncorrected	With control functions
Alternatives: 1. Antenna only, 2. Basic and extended cable, 3. Premium cable, 4. Satellite. Variables enter alternatives in parentheses and zero in other alternatives.		
	(Standard errors in parentheses)	
Price, in dollars per month (1-4)	-.0202 (.0047)	-.0969 (.0364)
Price for income group 2 (1-4)	.0149 (.0024)	.0150 (.0024)
Price for income group 3 (1-4)	.0246 (.0030)	.0247 (.0030)
Price for income group 4 (1-4)	.0269 (.0034)	.0269 (.0033)
Price for income group 5 (1-4)	.0308 (.0036)	.0308 (.0036)
Number of cable channels (2,3)	-.0023 (.0011)	.0026 (.0028)
Number of premium channels (3)	.0375 (.0163)	.0448 (.0235)
Number of over-the-air channels (1)	.0265 (.0090)	.0222 (.0110)
Whether pay per view is offered (2,3)	.4315 (.0666)	.5813 (.1089)
Indicator: ATT is cable company (2)	.1279 (.0946)	-.1949 (.1696)
Indicator: ATT is cable company (3)	.0993 (.1195)	-.2370 (.1760)
Indicator: Adelphia Comm is cable company (2)	.3304 (.1224)	.3425 (.1920)
Indicator: Adelphia Comm is cable company (3)	.2817 (.1511)	.2392 (.2250)
Indicator: Cablevision is cable company (2)	.6923 (.2243)	.1342 (.3381)
Indicator: Cablevision is cable company (3)	1.328 (.2448)	.7350 (.3521)
Indicator: Charter Comm is cable company (2)	.0279 (.1010)	-.0580 (.1377)
Indicator: Charter Comm is cable company (3)	-.0618 (.1310)	-.1757 (.1720)
Indicator: Comcast is cable company (2)	.2325 (.1107)	-.0938 (.2020)
Indicator: Comcast is cable company (3)	.5010 (.1325)	.1656 (.2227)
Indicator: Cox Comm is cable company (2)	.2907 (.1386)	-.0577 (.2300)
Indicator: Cox Comm is cable company (3)	.5258 (.1637)	.0874 (.2667)
Indicator: Time-Warner is cable company (2)	.1393 (.0974)	-.0817 (.1458)
Indicator: Time-Warner cable company (3)	.2294 (.1242)	-.0689 (.1803)
Education level of household (2)	-.0644 (.0220)	-.0619 (.0220)
Education level of household (3)	-.1137 (.0278)	-.1123 (.0278)
Education level of household (4)	-.1965 (.0369)	-.1967 (.0368)
Household size (2)	-.0494 (.0240)	-.0518 (.0241)
Household size (3)	.0160 (.0286)	.0134 (.0287)
Household size (4)	.0044 (.0357)	.0050 (.0358)
Household rents dwelling (2-3)	-.2471 (.0867)	-.2436 (.0865)
Household rents dwelling (4)	-.2129 (.1562)	-.2149 (.1562)
Single family dwelling (4)	.7622 (.1523)	.7649 (.1523)
Residual for extended-basic cable price (2)		.0805 (.0379)
Residual for premium cable price (4)		.0873 (.0380)
Alternative specific constant (2)	1.119 (.2668)	2.972 (.9176)
Alternative specific constant (3)	.1683 (.3158)	2.903 (1.301)
Alternative specific constant (4)	-.2213 (.4102)	4.218 (2.146)
Error components, standard deviation (2-4)	..5087 (.6789)	.5553 (.6410)
Log likelihood at convergence	-14660.84	-14635.47
Number of observations: 11810		

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 to \$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 TV reception: the education coefficients are progressively more highly negative for antenna-only (which is zero by normalization), extended-basic cable, premium cable, and 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. A dummy for whether the household rents its dwelling is included in the two cable alternatives and separately in the satellite alternative. These variables account for the fact 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 fact that it is relatively difficult to install a satellite dish on a multi-family dwelling. As expected, the estimated coefficient is positive.

The residuals from the first-stage price regressions enter the model to account for the omitted attributes. These control functions are created as follows. The price in each market was regressed against the product attributes listed in Table 1 plus Hausman (1997a)-type price instruments. The price instrument for market m is calculated as the average price in other markets that are served by the same multiple system operator as market m. A separate instrument is created for the price of extended-basic cable and the price of premium cable. Separate regressions were run for extended-basic price and premium price, using all of the instruments in each equation. The use of other instruments is discussed near the end of the paper in reference to both methods. The residuals were calculated from the estimated regressions. These residuals enter without transformation in the mixed logit model; that is, the control functions are the identity operator times a coefficient. The residual from the extended-basic cable price regression enters the extended-basic cable alternatives, and similarly for the premium cable. 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.

As stated above, the appropriate control function to include is a specification issue. We tried other specifications, including the use of both residuals in each alternative and the use of a series expansion (both signed and unsigned). They all provided nearly exactly the same results.

We turn now to the fixed effect approach. All of the elements of utility that do not vary within a market are subsumed into the fixed effects. The fixed effects are expressed as a function of price and other observed attributes:

$$\delta_{mj} = \alpha p_{mj} + \beta x_{mj} + \xi_{mj}.$$

The utility specification given above becomes:

$$U_{nj} = \delta_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} d_{gn} + k_j s_n + \sigma \nu_n c_j + \tilde{\epsilon}_{nj}.$$

Assuming $\tilde{\epsilon}_{nj}$ and ν_n are iid extreme value and standard normal respectively leads to a mixed logit of the same form as for the control function approach except with fixed effects for each alternative and market.

Estimation is performed in two stages. First the mixed logit model is estimated, using the contraction procedure described above for the fixed effects. Then the fixed effects are regressed against the product attributes using 3SLS. A separate equation is used for the extended-basic cable, premium cable, and satellite fixed effects, with the coefficients of the product attributes constrained across equations (so as to be consistent with the model in Table 1). The *negative* of the number of over-the-air channels enters these equations, since this attribute enters the antenna-only alternative in the model of Table 1 whereas it is now entering the fixed effects of the non-antenna alternatives.

The results are given in Table 2. 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 fixed effects on product attributes. The first column at the top gives the OLS results, which do not account for omitted attributes, and the second column gives the 3SLS results.

As with the control function approach, the correction for omitted variables raises the price coefficient. Without correction, three of the five in-

Table 2: Mixed Logit Model of TV Reception Choice
Fixed Effects Approach

Alternatives: 1. Antenna only, 2. Basic and extended cable, 3. Premium cable, 4. Satellite
Variable enters alternatives in parentheses and is zero in other modes.

Explanatory variable	OLS	3SLS
	(Standard errors in parentheses)	
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 (neg.) (1)	.0168 (.0132)	.0196 (.0186)
Whether pay per view is offered (2,3)	.5872 (.1326)	.7144 (.1814)
Indicator: ATT is cable company (2)	-.3458 (.2127)	-.2934 (.2353)
Indicator: ATT 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)
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 components, standard deviation (2-4)	.7775 (.1664)	
Log likelihood at convergence	-13927.40	
Number of observations: 11810		

come groups receive a positive estimated price coefficient. With correction, all groups obtain a significantly negative price coefficient.

The estimated base price coefficient is -.0922, compared to the -0.0969 obtained with the control function approach. The difference is not statistically significant at any reasonable confidence level. The estimates of θ_g, the incremental price coefficient for higher income groups, are very 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 becomes positive as expected when the endogeneity is corrected. All of the product attributes obtain similar values as with the control function approach. We tested the hypothesis that the coefficients of the product attributes and the base price coefficient are the same as the point estimates from the control function approach (i.e., as in Table 1.) The test statistic for a Wald test is 0.88, which with five degrees of freedom has a P-value of 0.9717, indicating that the hypothesis of equality cannot be rejected at any meaningful level of confidence. This test does not account for the variation in the estimates from the control function approach; however, the P value for a test that takes this variation into account would be even higher.

The demographic coefficients in Table 2 provide similar conclusions as those from the control function approach. Education induces households to buy less TV reception. Larger households tend not to buy extended-basic cable, and other differences are not significant. Renters tend not to buy cable and satellite as readily as owners. And single-family dwellers tend to purchase satellite reception more readily than households who live in multi-family dwellings.

Table 3 gives price elasticities from the models for each approach. Given that the price coefficients are nearly the same from the two methods, similar elastiticies would be expected, except for one issue. In particular, the two methods calculate elastiticies at different probabilities for each household. The fixed effects approach calculates elastiticies at the probabilities that arise when fixed effects are included in the model, such that forecasted shares equal sample shares in each market. For the control function approach, forecasted shares do not equal sample shares in each market, only for the sample as a whole.[4] We calculated elastiticies to determine whether this difference causes a difference in the elasticities. As the figures in Table 3 indicate, the two methods provide similar estimates.

[4] The model under the control function approach could be calibrated to each market prior to forecasting, such that market shares equal sample shares under this model also. This is a hybrid approach where fixed effects are not used in estimation but are calculated for forecasting.

Table 3:Estimated Elasticities

	Control Function	Fixed Effects
Price of extended-basic cable		
Antenna-only share	0.96	0.79
Extended-basic cable share	-1.18	-0.97
Premium cable share	0.99	0.88
Satellite share	0.95	0.87
Price of premium cable		
Antenna-only share	0.60	0.52
Extended-basic cable share	0.65	0.57
Premium cable share	-2.36	-2.04
Satellite share	0.64	0.58
Price of satellite		
Antenna-only share	0.43	0.42
Extended-basic cable share	0.48	0.43
Premium cable share	0.48	0.45
Satellite share	-3.79	-3.59

The elasticities under both approaches are smaller than those found by Goolsbee and Petrin on 1999 data. The difference might indicate that the cable elasticities have decreased over time as cable programming has become better and households have come to expect to watch the new cable shows.

As always with endogeneity, the selection of instruments is an issue. As stated above, we used the product attributes and Hausman-type prices as instruments, which follows the practice adopted in previous applications of the fixed effects approach, including the earlier work by Goolsbee and Petrin. To our knowledge, there has been no earlier application of the control function approach with cross-sectional market data; Villas-Boas and Winer use lagged prices in their time-series model. With disaggregate data for several markets, market-level averages of the demographic variables can serve as instruments. In aggregate models of demand and supply, the appropriate instruments include whatever demographic variables enter the aggregate demand function. In our model, the demand function is estimated on disaggregate data, such that the rationale for their inclusion as instruments for market-level price does not strictly apply. However, the underlying logic is still persuasive: the aggregate demographics can be expected to affect market price if disaggregate demographics affect customer-level choices. If the aggregate demographics are independent of the remaining error terms, then they can serve as useful instruments. We re-estimated the models with the demograpic variables included as extra instruments. The base price coefficient under the fixed effects approach dropped from -.0922 to -.0739, which, while noticeable, is not a statistically significant difference. With the control function approach, the base price coefficient was practically the same with or without the demographics as instruments. All the other coefficients were essentially unaffected in both approaches by use of these additional instruments. We tested the hypothesis that the coefficients of the demographic variables are collectively zero in the regressions of instruments on prices, and found that the hypothesis could not be rejected at any reasonable level of confidence. This result implies that the expected mean-squared error of the predicted price is higher when demographics are not included as instruments.

The use of Hausman-type price instruments is controversial (Bresnahan, 1997; Hausman, 1997b). 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 unobserved attributes. In aggregate demand-supply models, instruments are required beyond the observed variables that enter aggregate demand. These additional instruments are often difficult to obtain. The prices in other areas are useful because they are always available. With disaggregate demand models, the need for additional instruments is not as stringent. In particular, aggregate demographics do not enter the disaggregate models but affect market price. They can serve as the extra instruments that are needed for demand estimation, providing price variation over households that is not contaminated by differences in omitted attributes and yet not collinear with the other variables that enter the demand model.[5] It is possible therefore for the instruments to consist of only the observed product attributes and aggregate demographics. (Extra instruments are of course useful; the point here is that they are not required.) As a result, the strong motivation for using the Hausman-type instruments that arises with aggregate models is not present with disaggregate demand data. We re-estimated the model without using the prices in other areas as instruments but, as in the previous paragraph, including the aggregate demographics (since instruments beyond the product attributes are required.) With the control function approach, the estimated price coefficient rose when the Hausman-type prices were removed as instruments. This is the direction of change that would be expected if

[5] Consider two households that have the same demographics but live in areas where 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.

the prices in other markets incorporated the impact of unobserved attributes that were correlated over markets. However, this change did not occur in the fixed effect approach: removing the Hausman-type prices as instruments had essentially no effect on the base price coefficient. The other coefficients were not affected under either method.

It perhaps useful to compare the methods from a programming perspective. Both procedures are fairly easy to implement. The control variable approach can be implemented with available software packages such as SAS (which now has a mixed logit and probit routine), LIMDEP, and the codes available from Train's website. The fixed effects approach can be applied with these packages if the number of fixed effects is sufficiently small that they can be treated as regular parameters in the estimation routines. With numerous fixed effects, the contraction procedure is needed. However, this procedure is easy to implement in programming languages; the version of Train's mixed logit code that he adapted for the fixed effects models in this paper is available from him on request. The codes for fixed effects run just about as fast as those for the control function approach; the difference in run times is not sufficient to serve as the basis for choosing one over the other.

Appendix

We obtained information on households' television choices, the characteristics of households, and the prices and attributes of the cable franchise serving the household's geographic area. This information comes from two sources, the Forrester Technographics 2001 survey and Warren Publishing's 2001 Television and Cable Factbook. The Forrester survey was designed to be a nationally representative sample of households. It asks respondents about their ownership and use of various electronic and computer-related goods. To these data we match information about cable franchises from Warren Publishing's 2001 Factbook, which is the most comprehensive reference for cable system attributes and prices in the industry.

To minimize sampling error in market shares, we restricted our analysis to markets where there are at least 30 respondents in the Forrester survey. This screen yields 300 cable franchise markets with a total of almost 30,000 households. We randomly choose 172 of these 300 markets, so as to reduce the number of fixed effects that needed to be estimated. From these 172 markets, we randomly selected 11810 households, oversampling those households from smaller markets (again, to minimize sampling error). These 11810 households are used in the estimation with weights equal to the inverse of their probability of being sampled.

As stated in the body of the paper, the alternatives in the discrete choice model are: expanded basic cable, premium cable (which can only be purchased bundled with expanded basic), Direct Broadcast Satellite, and no multi-channel video (i.e., local antenna reception only). In the Forrester survey, respondents reported whether they have cable or satellite, and the amount they spend on premium television. We classified respondents as having premium if they reported that they have cable and spend more than $10 per month on premium viewing, which is the average price of the most popular premium channel, HBO. We classified respondents as choosing expanded basic if they reported that they have cable and they spend less than $10 per month on premium viewing.

The survey provides various demographic characteristics, including family income, household size, education, and type of living accommodations. It also includes an identifier for the household's television market, which can be used to link households to their cable franchise provider.

The cable franchise market of each surveyed household was matched to cable system information from Warren Publishing's 2001 Television and Cable Factbook. The attributes we include are the channel capacity of a cable system, the number of pay channels available, whether pay per view is available from that cable franchise, the price of basic plus expanded basic service, and the price of premium service. We also obtain from the Factbook the number of over-the-air channels available in the franchise market. Finally, for the price of satellite, we use \$50 per month plus an annual \$100 installation and equipment cost.

References

Berry, S. (1994), 'Estimating discrete choice models of product differentiation', *RAND Journal of Economics* **25**, 242–262.

Berry, S., J. Levinsohn and A. Pakes (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**, 841–889.

Berry, S., J. Levinsohn and A. Pakes (2001), 'Differentiated products demand system from a combination of micro and macro data: The new car market', Working Paper, National Bureau of Economic Research.

Bhat, C. (2001), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model', *Transportation Research B* **35**, 677–693.

Blundell, R. and J. Powell (2001), 'Endogeneity in semiparametric binary response models', Working Paper, Department of Economics, University College London.

Bresnahan, T. (1997), 'The apple-cinnamon cheerios war: Valuing new goods, identifying market power, and economic measurement', Working Paper, Stanford University.

Brownstone, D. and K. Train (1999), 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–129.

Crawford, G. (2000), 'The impact of the 1992 cable act on household demand and welfare', *RAND Journal of Economics* **31**, 422–449.

Goolsbee, A. and A. Petrin (2002), 'The consumer gains from direct broadcast satellites and the competition with cable tv', Working Paper, Graduate School of Business, University of Chicago.

Halton, J. (1960), 'On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional integrals', *Numerische Mathematik* **2**, 84–90.

Hausman, J. (1978), 'Specification tests in econometrics', *Econometrica* **46**, 1251–1272.

Hausman, J. (1997*a*), 'Reply to professor bresnahan', Working Paper, MIT.

Hausman, J. (1997*b*), Valuation of new goods under perfect and imperfect competition, *in* R.Gordon and T.Bresnahan, eds, 'The Economics of New Goods', University of Chicago Press, Chicago.

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

Heckman, J. and R. Robb (1985), 'Alternative methods for evaluating the impacts of interventions: An overview', *Journal of Econometrics* **30**, 239–267.

Hensher, D. (2001), 'The valuation of commter travel time savings for car drivers in new zealand: Evaluating alternative model specifications', *Transportation* **28**, 101–118.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P.Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York, pp. 105–42.

Munizaga, M. and R. Alvarez-Daziano (2001), 'Mixed logit versus nested logit and probit', Working Paper, Departmento de Ingeniera Civil, Universidad de Chile.

Nevo, A. (2001), 'Measuring market power in the ready-to-eat cereal industry', *Econometrica* **69**, 307–342.

Petrin, A. (2002), 'Quantifying the benefits of new products; the case of the minivan', *Journal of Political Economy* **110**, 705–729.

Smith, R. and R. Blundell (1986), 'An exogeneity test for a simultaneous equation tobit model with an application to labor supply', *Econometrica* **54**, 679–686.

Train, K. (1986), *Qualitative Choice Analysis*, MIT Press, Cambridge, MA.

Train, K. (1998), 'Recreation demand models with taste variation', *Land Economics* **74**, 230–239.

Train, K. (2000), 'Halton sequences for mixed logit', Working Paper No. E00-278, Department of Economics, University of California, Berkeley.

Train, K. (2002), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Villas-Boas, J. and R. Winer (1999), 'Endogeneity in brand choice models', *Management Science* **45**, 1324–1338.