Omitted Product Attributes in Differentiated Product Models

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

April 25, 2004

Abstract

Omitted attributes in differentiated product models can generate a severe downward bias in price elasticities. We develop a control function method for this problem. It includes proxies for unobserved characteristics that are functions of the differences between product prices and their predicted values conditional on all demand and supply factors observed by the econometrician. The main alternative is to use product-market controls, which protects against a wider class of omitted variable problems, but requires the estimation of many additional parameters in a non-linear system of equations, and is not available in all settings. We benchmark results from an uncorrected approach and our control function method against the product-market control approach for three data sets and demand specifications estimated elsewhere that span a range of markets, levels of aggregation, and potential competitive pricing behaviors, including automobiles (the original Berry, Levinsohn, and Pakes (1995) application), cable television, and margarine. The estimated elasticities are very similar across the control function and product-market control approaches, and they both differ significantly from the uncorrected elasticity estimates, which are severely biased down in every case.

1 Introduction

Models of differentiated products are widely used for estimating demand elasticities and substitution patterns, which are a critical input into policy analysis, questions about the design of new products, and a host of other issues.[1] In applications of these models it is rare that all relevant product attributes are observed by the econometrician. When some attributes are omitted, price will typically be correlated with the unobserved portion of consumers' utility for that product: the equilibrating mechanism in the market causes the price to be higher for products that display desirable attributes observed by consumers and producers but not measured by the econometrician. This problem arises for both aggregate (i.e. market-level) data and disaggregate (i.e., customer-level) data, and the positive correlation between price and the unobserved attribute biases estimates of price elasticities towards zero.

Since the demand for differentiated products under heterogeneous preferences is inherently non-linear, standard instrumental variable methods for correcting this endogeneity problem are not immediately applicable. Including product-market fixed effects directly in the specification for utility controls for the unobserved factors.[2] However, this method is often computationally intractable; the objective function is non-linear and not generally globally concave in parameters, leading to a curse of dimensionality as the function must be maximized with the hundreds or thousands of additional fixed effect parameters.

Berry (1994) and Berry, Levinsohn, and Pakes (1995, henceforth BLP) develop an alternative approach, which is similar in spirit to a fixed effects approach. Berry proves the existence and uniqueness of a set product-market controls that match observed to predicted market shares. BLP provide a method for locating them, conditional on the other model parameters. The method has been used successfully in numerous applications with aggre-

[1] Examples where measuring substitution patterns is important include: forecasting whether the induced demand for new energy-efficient vehicles is drawn more from "gas-guzzlers" or current "gas-sippers", estimating the price effects and welfare implications of a merger, and understanding whether a new product will lead to self-cannibalization of the firm's sales of its other products.

[2] A second stage of estimation is often required if parameters that are not separately identified from the fixed effects are necessary for the analysis or forecasts.

gate data, including the demand for cable TV (Crawford (2000)), cereals (Nevo (2001)), and minivans (Petrin (2002)), to name only a few. The BLP approach can also be applied to disaggregate data, or a combination of aggregate and disaggregate data, as illustrated by Berry et al. (2004) and Goolsbee and Petrin (2004).

While applicable in many situations, the approach is not fully general. The BLP estimator is not consistent and asymptotically normal (CAN) if the magnitude of the sampling error in the market shares is not small, because the error does not average out of the estimating equations.[3] BLP is not available if the goods are not all substitutes, which is required in Berry's proof of uniqueness for the product-market controls. Finally, the BLP contraction algorithm that locates the controls requires (for convergence to be guaranteed) that an independent and identically distributed (i.i.d.) logit error be appended to the utility function. Thus, the BLP estimator is inconsistent if the i.i.d. logit error results in a misspecified utility function, as shown in a number of empirical applications using both consumer-level and market-level data, including Petrin (2002), Goolsbee and Petrin (2004), and Song (2003).[4]

We develop an alternative approach for omitted attributes that is based on control functions, which are extra variables that are added to the estimation equation to control for the part of the error that is correlated with the regressors, as in Heckman (1978) and Hausman (1978).[5] In the context of differentiated products, where unobserved product attributes lead to endogenous prices, the idea is to add to the demand side model proxies for these unobserved demand characteristics.

[3]Berry, Linton, and Pakes (2003) report that the number of observed purchasers must grow like J^2, where J is the number of products.

[4]Note that, in this last case, the product-market controls exist and are unique for specifications without the logit error, but are so computationally difficult to locate that the BLP procedure is effectively unavailable, as in the fixed effects case described earlier. Goolsbee and Petrin (2004) estimate a model with product-market controls and no logit error, but are able to do so only because they have 4 products per market, so solving the non-linear system of equations that matches observed to predicted market shares market-by-market without using the contraction is computationally tractable.

[5]The first use of the term "control function" of which we are aware is in Heckman and Robb (1985) in the context of selection models. It has been applied to a Tobit model by Smith and Blundell (1986) and binary probit by Rivers and Vuong (1988). See also Blundell and Powell (2001).

As in Trajtenberg (1989,1990), Villas Boas and Winer (1999), and Bajari and Benkard (2003), we exploit the information that prices contain on unobserved demand factors. Our control function approach is novel in the sense that we write equilibrium prices as a function of the demand and cost characteristics of all goods and the identities of their producers in the market. The control function has as arguments the differences between product prices and their predicted values given all of the relevant demand and supply factors that the econometrician observes. We develop the conditions under which these differences can provide a basis for the unobserved demand factors, so they can be proxied out of the estimation equations.

The control function approach we outline is straightforward to implement in standard programming packages; the first stage is a regression, and the second stage is maximization of a likelihood function. While the approach is not as robust to omitted attributes as including a control for each product in every market, it is consistent for some well-known and pragmatic pricing rules. Even in the cases where the basis does not fully span the space of unobserved demand attributes, the approach still provides an easy-to-implement test that will often have high power in the case when price endogeneity is a problem, and the estimates are always useful as starting values for any of the more demanding approaches (like fixed effects or BLP). It applies to choice sets that include all types of goods (that is, not just choice sets in which goods are all substitutes for one another). It does not match observed to predicted market shares in the estimation routine, so sampling error in market shares is not a primary concern. This also obviates the need for the BLP contraction algorithm, which simplifies computation and alleviates the need for the i.i.d. logit error, avoiding any specification problems that may arise with it.

We present three empirical demand applications that replicate specifications from earlier works - all of which use product-market controls - including Berry, Levinsohn, and Pakes (1995), who look at automobiles, Goolsbee and Petrin (2004), who look at cable television, and Chintagunta, Dube, and Goh (2003), who look at margarine. These applications are chosen to span a range of markets and potential competitive pricing behaviors, and include three types of different data: aggregate (market-level) data, household-level cross-sectional data, and household-level panel data. We use them to show in practice how one implements the control function approach for these different types of products and aggregation levels. We test for the presence of price endogeneity by comparing the control function approach to a standard uncorrected approach and reject the null of no price endogeneity in every case. The estimated elasticities are almost identical across the BLP and control function approaches and differ significantly from the uncorrected elasticity estimates, which are severely biased down in every case.

The paper proceeds as follows. Section 2 describes differentiated products demand models and the endogeneity problem. Section 3 describes the control function approach. Section 4 describes the benchmark - product-market controls - against which we compare the control function results. Section 5 compares the approaches. Section 6 reports the results from the three applications using real data. Section 7 explores a number of Monte Carlo specifications suggested to us by Wolak (2003), and Section 8 concludes.

2 Demand and Endogenous Prices

The problem of omitted attributes and endogenous prices arises naturally within characteristics' based demand approximations. At the core of these approaches is a desire for more parsimony than the unrestricted constant-elasticity-of-demand system (for J goods there are J^2 parameters). Characteristics' based approaches "solve" this problem by assuming that demands for J goods can be reasonably approximated by $K << J$ characteristics, where the K factors serve as the basis for utility.[6] The approximation can be quite parsimonious, in some cases reducing the number of necessary parameters sufficient to infer all own- and cross-price elasticities to K.[7]

When the K factors used by the econometrician exclude important characteristics, price can be endogenous. To illustrate this phenomenon, we use the formulation of utility given in Berry (1994) and BLP:

$$u_{ij} = x_j\beta_i - \alpha_i p_j + e_{ij}, \tag{1}$$

[6] Other assumptions include: utility is only derived from the characteristics of the chosen good, for any given consumer tastes for these characteristics are constant across goods, and (typically) a distributional assumption on the error in the latent variable (utility) model for demand. See e.g. McFadden (1981).

[7] In most applications, K is usually less than 10 and $J^2(J)$ can be as large as 40,000 (200).

where x_j and p_j are the observed (by the econometrician) product characteristics and price, and β_i and α_i are consumer-specific tastes that can be a function of both demographics and idiosyncratic tastes. e_{ij} is the two-component "error" entering utility, given as

$$e_{ij} = \xi_j + \epsilon_{ij}.$$

ξ_j is product-specific, capturing characteristics known to both consumers and producers in the market, but not included in the estimated specification. ϵ_{ij} is a consumer-specific idiosyncratic taste that is assumed to be independent across both consumers and products. Letting e_i denote the vector $\langle e_{i1}, \ldots, e_{iJ} \rangle$, and $d\varphi(e_i)$ the density of e_i conditional on the observed variables, the choice probability for good l given β_i and α_i is

$$P_{il} = \int_{A_{il}} d\varphi(e_i) \tag{2}$$

where $A_{il} = \{e_i \mid U_{il} > U_{ij} \; \forall j \neq l\}$ is the set of e_i such that product l provides maximal utility.

To focus on price determination, index markets by m and define z_{mj} as the set of the observed characteristics that affect demand (x_{mj}) and marginal costs (w_{mj}), so all observed product characteristics in market m are included in the vector $z_m = (z_{m1}, z_{m2}, \ldots, z_{mN_m})$, where N_m is the number of goods in market m. A characteristic here includes indicators for the seller of each product, as oligopoly models are explicit in their recognition that equilibrium prices depend on the pattern of product ownership when firms sell multiple products. Unobserved characteristics for product j in market m are given by (ξ_{mj}, ω_{mj}), where ω_{mj} is an unobserved cost factor. d_m denotes market-level demographics. As long as regularity conditions hold such that prices can be expressed as an implicit function of all exogenous factors, price for good j can be written as

$$p_{mj} = p_j(z_m, \xi_m, \omega_m, d_m); \tag{3}$$

it will generally have as arguments all of the factors $(z_m, \xi_m, \omega_m, d_m)$ that are taken to be given at the time that firms make pricing decisions.[8]

[8] These assumptions are standard in the literature.

The econometric problem occurs because e_{ij} and p_j are not typically independent, as is often assumed in standard choice models.[9] The dependence arises because e_{ij} in part reflects the utility generated by unmeasured characteristics (via ξ_j) common to all consumers. Sellers will typically charge higher prices when their products have more desirable *omitted* characteristics, leaving e_{ij} and p_j correlated even after conditioning on x_j. When independence is maintained, consumers look less price sensitive than they actually are because they are getting more for paying the price they pay than the econometrician has taken into account.

Since Trajtenberg (1989)'s finding of upward sloping demand curves for CAT scanners, numerous empirical applications have shown that e_{ij} and p_j can be so highly correlated in practice as to preclude the use of the characteristics approach entirely. Other examples where the presence of this correlation is important empirically include automobiles (BLP and Petrin (2002)), cable television choices (Goolsbee and Petrin (2004) and Crawford (2000)), supermarket goods like cereals (Nevo (2001)), yoghurt and ketchup (Villas-Boas and Winer (1999)), and margarine and orange juice (Chintagunta, Dubé, and Goh (2003)), to name a few.

3 Control Functions for Omitted Attributes

Control function approaches include an additional proxy in the estimating equation to test for and/or correct an omitted variables problem. In the context of differentiated products, the idea is to add to consumers' utility functions terms that proxy for the unobserved product characteristics. We develop a control function method for differentiated products in this section and show how to apply it for different data types and markets in Section 6.[10]

We outline the method first. It consists of two steps. In the first step

[9]Logit and GEV models (e.g., McFadden (1974, 1978)) assume that the unobserved component of utility is independent of the observed variables. Mixed logit and probit (e.g., Brownstone and Train (1999)) allow the covariance of the unobserved component to depend on observed variables; however, the mean is assumed to be constant, which precludes correlation with price.

[10]STATA code for implementing the control function approach with random coefficients will soon be made available at http://gsbwww.uchicago.edu/fac/amil.petrin/research/. Contact the authors directly for preliminary code.

regression is used to estimate the variables that enter the proxy function. In the second step the likelihood function is maximized with the proxy functions added as extra explanatory variables. Specifically, for each product j the new variable that enters every consumer's utility function - the control function for j - is denoted $f_j(\mu_m;\lambda)$, where $\mu_m = \langle \mu_{m1}, \ldots, \mu_{mJ} \rangle$ is the vector of variables obtained from the first stage regressions and λ is a vector of parameters. Utility is rewritten as

$$u_{ij} = \alpha_i p_{mj} + \beta_i x_{mj} + f_j(\mu_m;\lambda) + \eta_{mj} + \epsilon_{ij}, \tag{4}$$

which is the original formulation from (1) with the control function added to it and the residual

$$\eta_{mj} = (\xi_{mj} - f_j(\mu_m;\lambda))$$

representing the difference between the control and the actual value of ξ_{mj}. (4) illustrates that the important specification questions for the control function approach involve the choice of μ_m, the choice of the proxy $f_j(\mu_m;\lambda)$, and the specification for the new error component η_{mj}.

As in Trajtenberg (1989,1990), Villas Boas and Winer (1999), and Bajari and Benkard (2003), the idea here is to exploit the information that prices contain on unobserved demand factors. We write the equilibrium pricing function as $p_{mj} = p_j(z_m, \xi_m, \omega_m, d_m)$, with prices potentially determined by all observed and unobserved exogenous factors in the market, including the identity of products' producers.[11] We explore a set of restrictions on this function and the data generating process that allow information on ξ_m to be recovered.

We start with the assumption that the unobserved factors (ξ_m, ω_m) are independent of the observed factors (z_m, d_m), a strong but maintained primitive in this literature.[12] Given the independence, we ask what one can learn about ξ_m if prices are additively separable in these observed and unobserved

[11] The same set of products divided up differently across the set of multi-product producers may result in different equilibrium prices as producers internalize pricing externalities.

[12] For example, BLP maintain that the unobserved factors are mean independent of observed factors. The literature has attempted to address what is likely to be the more severe problem of the unobserved characteristic's correlation with price, the most easily adjusted characteristic.

factors. p_{mj} can then be written as

$$\begin{aligned} p_{mj} &= g_{1j}(z_m, d_m) + g_{2j}(\xi_m, \omega_m) \\ &= E[p_j \,|\, z_m, d_m] + \mu_j(\xi_m, \omega_m), \end{aligned} \tag{5}$$

where in the second line we define the difference between price and its expected value conditional on observed exogenous factors as $\mu_{mj} = \mu_j(\xi_m, \omega_m)$, an argument in the control functions.[13]

If the data generating process is one in which the number of observed markets is large relative to the number of products, variation in prices and observed characteristics and the mean independence assumption allow one to consistently estimate $E[p_j \,|\, z_m, d_m]$ and $\mu_j(\xi_m, \omega_m)$. This is the setting for both the cable television and margarine cases, where there are four product types, and variants of them are observed in a cross-section (U.S. cable franchise markets in 2001) and a time-series (margarine sales at a supermarket over 117 weeks).

When the number of arguments entering $E[p_j \,|\, z_m, d_m]$ is large relative to the number of observed markets, a dimensionality problem can arise. The first order approximation to the unrestricted basis for this expectation in a market with J products has dimension JK (at least), where K is the number of observed characteristics per product. For this situation we suggest using an approximation to $E[p_j \,|\, z_m, d_m]$ suggested in Pakes (1994). He provides a parsimonious basis for equilibrium pricing functions that are partially exchangeable (that is, when one can change the order in which some of the

[13]Villas-Boas and Winer use last period's price as the argument in the expected price function. Trajtenberg uses only own-product characteristics (no exogenous factors except those already entering the own-product utility function). He also does not enter the actual price, and thus does not estimate the price coefficient along with the rest of the model estimates. Instead he uses additional outside information to estimate the price coefficient (as he describes on pp. 463-465 of the 1989 paper).

Bajari and Benkard (2003)) show that when product space is "full," that is, where every set of product characteristics is produced and consumed (as in Rosen (1974)), and where many products with the same observed characteristics but different unobserved characteristics exist, the generality given by the pricing functions in (3) is unnecessary. In particular, the fullness of product space ensures the market is competitive, so the supply side can be ignored, and price can be written as $p_{mj} = p(x_{mj}, \xi_{mj})$, a function of *only* the own-product observed and unobserved demand characteristics. They then infer proxies for the unobserved demand factor by comparing the prices of products with identical observed demand characteristics; the bigger the difference, the larger the inferred unobserved demand factor.

arguments enter the function without changing the value of the function). In particular, he shows that the first order terms for the basis are of dimension 3K. This result has wide applicability to differentiated product markets, and we discuss its usefulness further in the context of the control function approach in the BLP automobile case-study.

With estimates of $\mu_j(\xi_m, \omega_m)$, the question arises as to the functional form for $f_j(\mu_m; \lambda)$. In principle, every product residual μ_{mj} contains information on every ξ_{mj} (although most of the explanatory power for ξ_{mj} is likely to come from the regressor μ_{mj}). Thus, the entire set of residuals μ_m should be used as a basis for each unobserved factor. For example, a first order approximation would be given by $f_j(\mu_m; \lambda) = \overline{\lambda}_j' \mu_m$, where $\overline{\lambda}_j = \langle \lambda_{j1}, \ldots, \lambda_{jJ} \rangle$.[14] If there are lots of products in each market, this can again raise a dimensionality problem.[15] Arguments from Pakes (1994) can be used to develop a more parsimonious basis for $f_j(\mu_m; \lambda)$, as we also illustrate in the automobile example.

A final issue concerns the presence of the unobserved cost shocks in the price residuals. Conditional on ξ_{mj}, the cost shocks ω_m are not correlated with any of the other variables entering utility for product j (this is also true for ξ_{mk} $\forall k \neq j$). This means that ω_m enters only as an additional source of noise in the basis for $f_j(\cdot)$. Note that the more highly correlated the cost shocks ω_m are with the demand shocks ξ_m, the less noise this source of error is likely to add. Since profit maximizing firms avoid incurring costs unless they are associated with new attributes for which consumers are willing to pay, demand and cost shocks may be reasonably correlated in applications.

The extent to which $f_j(\mu_m; \lambda)$ proxies for ξ_{mj} in any particular case is an empirical question. As is clear, there are two possible sources of error in this approach. First, since the first stage pricing equation is estimated, $\hat{\mu}_m$ is used in the basis instead of μ_m.[16] Second, the presence of ω_m in μ_m can also add noise to the basis. For each product j we suggest approximating the error in the proxy $\eta_{mj} = (\xi_{mj} - f_j(\mu_m; \lambda))$ by adding a random effect to the estimated specification.

In some cases, the choice for $f_j(\mu_m; \lambda)$ and the distribution of η_{mj} is

[14] One can add higher order terms to this function, or non-parametrically approximate it with enough degrees of freedom.

[15] For the first order basis, for J products there are J^2 parameters.

[16] In the applications we describe how to account for this source of error in the standard errors of the parameter estimates.

determined exactly by the underlying primitives. For example, if ξ_m and μ_m are distributed multivariate normal, the conditional mean of ξ_m given μ_m is given as $\Omega_{\xi\mu}\Omega_{\mu\mu}^{-1}\mu_m$, where $\Omega_{\xi\mu}$ is the covariance between the demand errors and the price disturbances and $\Omega_{\mu\mu}$ is the variance covariance matrix for the price disturbances. Control functions are then linear in the price residuals, so $f_j(\mu_m;\lambda) = \overline{\lambda}_j'\mu_m$, where $\overline{\lambda}_j = \langle \lambda_{j1}, \ldots, \lambda_{jJ} \rangle$, and $\eta_{mj} = (\xi_{mj} - \overline{\lambda}_j'\mu_m)$ is normally distributed, leading to:

$$u_{ij} = \alpha_i p_{mj} + \beta_i x_{mj} + \overline{\lambda}_j'\mu_m + \eta_{mj} + \epsilon_{ij}. \tag{6}$$

In the case where the covariances across products are zero, so $\Omega_{\xi\mu}$ and $\Omega_{\mu\mu}$ are both diagonal, the control function for each product is simply proportional to the price residual for that product, so $f_j(\mu_m;\lambda) = \lambda_j\mu_{mj}$, and again η_{mj} is normally distributed:[17]

$$u_{ij} = \alpha_i p_{mj} + \beta_i x_{mj} + \lambda_j\mu_{mj} + \eta_{mj} + \epsilon_{ij}. \tag{7}$$

Cost-plus pricing - which Shim and Sudit (1995) find is used by over 80% of managers at manufacturing firms - provides one set of examples where the control function approach is exact.[18] These rules-of-thumb set prices by using percentage markups m_j over costs mc_j, as in the case when price is set as $p_{mj} = (1 + m_j) * mc_{mj}$.[19] Let marginal costs be given by $mc_{mj} = \gamma_0 + \gamma' w_{mj} + \omega_{mj}$, with γ_0 and γ respectively a scalar (intercept) and vector of cost function parameters. If demand and cost shocks are distributed joint normal, ω_{mj} can be expressed as $\omega_{mj} = \tau_0 + \tau_1\xi_{mj} + \iota_{mj}$,

[17]This is the formulation used in Villas-Boas and Winer (1999).

[18]They report survey results from 141 manufacturing firms spanning: small (less than $10 million in sales) to large ($1 to $5 billion in sales), single and multi-product firms, and industries such as chemicals, machining, electronics, transportation, and medical. 82% of these firms use one of these rules as the method for determining prices, with their use being fairly uniform across all of these groups. They suggest that managers find it too costly to fully determine the marginal revenue curve in a timely manner given the many goods which they are pricing at these multi-product firms.

[19]These methods go by a variety of terms (cost plus pricing, markup pricing, full cost pricing, and variable cost pricing, to name a few), and are the leading examples of pricing methods taught in marketing textbooks. See, for example, the textbook on marketing by Kotler (2000), where numerous cost plus techniques are discussed. These pricing methods are also employed by academics when modeling pricing behavior (see, e.g., Yang, Chen, and Allenby (2003)).

with ι_{mj} normally distributed. In this case

$$\begin{aligned} p_{mj} &= (1+m_j)*(\gamma_0+\gamma' w_{mj}+\omega_{mj}) \\ &= (1+m_j)*(\gamma_0+\tau_0+\gamma' w_{mj})+(1+m_j)*(\tau_1\xi_{mj}+\iota_{mj}), \end{aligned}$$

so the difference between p_{mj} and $E[p \mid j, z_m, d_m]$ is given by $\mu_{mj}=(1+m_j)*(\tau_1\xi_{mj}+\iota_{mj})$, which is distributed joint normal with ξ_{mj}. The result extends to the case when ξ_m and ω_m are distributed joint normal, which results in ξ_m and μ_m being distributed joint normal (in this case, the control function is given as $f_j(\mu_m;\lambda)=\overline{\lambda}_j'\mu_m$).[20]

In general, the difference between the true distribution for η_{mj} and the distribution that is used to approximate it determines the extent of the measurement error. Corrections for measurement error in a non-linear framework have been the source of much attention recently (see, for example, Carroll, Ruppert, and Stefanski (1995) and Hong and Tamer (2002)).[21] We leave for future work the application of these types of corrections.

Even in the cases when $f_j(\mu_m;\lambda)$ measures ξ_{mj} with error (and the chosen specification for the random effect does not correctly parameterize this error), the presence of $f_j(\mu_m;\lambda)$ in the estimation will typically work to condition out some of the unobserved demand attribute. For this rea-

[20]It is straightforward to show that the example extends to a pricing rule given by $p_{mj}=mc_{mj}+m_j$. The example also carries over to more general cases for marginal cost. For example, the control function approach will work when marginal costs are log-linear in characteristics, or $mc_{mj}=e^{\gamma_0+\gamma' w_{mj}+\omega_{mj}}$, so price is given by (for example)

$$p_{mj} = (1+m_j)*e^{\gamma_0+\gamma' w_{mj}+\omega_{mj}}.$$

In this case the log of price minus the expectation of the log of price conditional on observed exogenous factors is the appropriate difference for the control function residual, because

$$ln(p_{mj})-(ln(1+m_j)+\gamma_0+\tau_0)-\gamma' w_{mj}=\tau_1\xi_{mj}+\iota_{mj},$$

which is distributed joint normal with ξ_{mj}. Thus, to the extent that marginal costs may be multiplicative, the difference in log of price and its expected value is a natural second specification to check for robustness of the control function results.

[21]For example, in Hong and Tamer(2002), they show that in a non-linear method of moments model with Laplace distributed measurement errors, it is possible to derive revised moment conditions that are in terms of the observed variables and that can be used to make consistent inference about the model parameters. As is well-known, the first order conditions from a the maximum likelihood estimator give maximum likelihood a non-linear method of moments characterization.

son, a comparison of the coefficients from specifications estimated with and without the control function can provide a powerful test for the existence of price endogeneity induced by omitted attributes, even when the control function approach may be inconsistent.[22] Thus, before undertaking a more complicated approach, the control function approach is always available for helping a researcher determine whether the estimated elasticities from the uncorrected model are appropriate for policy analysis. The control function estimates are then available as starting values for these more involved approaches, which we now turn to discussing.

4 The Benchmark: Product-Market Controls

Another way to address the price endogeneity problem is to include some form of alternative-specific controls for each product-market pair (like fixed effects). Berry (1994) and Berry, Levinsohn, and Pakes (1995) have shown the applicability of this approach to discrete choice demand models. The main advantage of this approach is that it allows for a very general error structure at the product-market level that is robust to many forms of correlation between price and the unobserved attribute. For this reason, in the empirical applications we use a model with product-market controls as a benchmark for determining how well the control function approach performs.

To show how the product-market controls protect against endogenous prices, BLP decompose the original utility specification from (1) into three components:

$$u_{ij} = \delta_{mj} + v_{ij} + \epsilon_{ij}. \tag{8}$$

δ_{mj} is the fixed component of utility - the same for all consumers - and is given by

$$\delta_{mj} = \alpha_0 p_{mj} + \beta_0' x_{mj} + \xi_{mj}. \tag{9}$$

It collects all the terms in utility that do not vary beyond the product-market pair, including the omitted attribute ξ_{mj}. v_{ij} represents consumer-specific taste variation that is attributable to idiosyncratic tastes for characteristics

[22]This general idea is one of the proposed uses of the control function approach in non-linear environments (see Smith and Blundell (1986)), although the context there is different.

or tastes that derive from demographics:

$$v_{ij} = (\alpha_i - \alpha_0)p_{mj} + (\beta_i - \beta_0)'x_{mj}. \tag{10}$$

Finally, ϵ_{ij} is, as before, an idiosyncratic product-consumer taste shock that is i.i.d. across both products and consumers.

The approach consists of two stages of estimation. First, the discrete choice model is estimated with alternative-specific constants for each product in each market. Parameters not captured by these δ_{mj}'s are consistently estimated because ξ_{mj} is controlled for (the error is ϵ_{ij}, which is not correlated with price). A second stage of estimation must be undertaken to recover estimates of the parameters included in the δ_{mj}'s.

The second stage uses the δ_{mj}'s as the dependent variable in an instrumental variables regression of (9). While the x_{mj} are exogenous (by assumption), price is endogenous, that is, it is correlated with ξ_{mj}, and is instrumented. With estimates for all parameters available after this last step, any function of the parameters (like price elasticities) is straightforward to compute.

5 Comparison of Approaches

Both approaches are designed to test for omitted characteristics that are correlated with price and to provide a correction for them. Both approaches depend fundamentally on the maintained assumption of mean independence of the unobserved factors. They also depend on the availability of variables that enter the equilibrium pricing function for $p_j(\cdot)$ but that do not enter utility for product j, conditional on the observed characteristics for the good. Our control function approach assumes that equilibrium prices are additively separable in observed and unobserved characteristics. Including a control for every product-market pair provides for an estimator that alleviates the need for this separability assumption.

Fixed effects are perhaps the most natural set of product-market controls to use. The biggest hurdle facing a researcher using fixed effects arises because these controls are not easily concentrated out of the objective function in a non-linear environment. Since the objective function is almost never globally concave, when there are lots of products and/or markets, the number of additional parameters can make the estimator intractable. The automobile market, for example, from 1971-1990 (the BLP empirical case study) requires over 2200 product-market controls. Without global concavity, no foreseeable gain in computing speed will allow for the maximization of an objective function with hundreds or thousands of parameters entering non-linearly.

Berry (1994) develops an alternative approach that is similar in spirit to a fixed effects model. For a general class of choice models he shows the existence and uniqueness of a set of product-market indicators that induce forecasted shares to equal actual shares in every market. However, except for some special cases outlined by Berry, where data is aggregated to the market-level and the error structure is i.i.d. logit or nested logit, these product-market controls raise the same issue as the fixed effects estimator; they enter the objective function non-linearly and computationally can be difficult to concentrate out.

BLP develop one solution to this problem. They provide an algorithm that solves for the unique vector of product-market indicators that match observed to predicted market shares, conditional on the other parameters in the model.[23] They prove that this algorithm is guaranteed to converge (i.e. it is a contraction). These insights of Berry and BLP together are the key to implementing a model with product-market controls.

While applicable in many situations, BLP is not completely general. Because it matches observed to predicted market shares to invert out the unobserved characteristics, sampling error enters the estimation equations in a non-linear manner. For this reason, unless sampling error in the market shares is minimal, the BLP estimator is not consistent and asymptotically normal (CAN). Specifically, the number of observed purchasers must grow at the rate J^2 (where J denotes the number of products), a demanding requirement for some data generating processes.

A second problem can arise if one needs to use the contraction to concentrate out the product-market controls (as is almost always done in prac-

[23]The algorithm calculates the product-market controls 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} $\forall j$. As mentioned above, this iteration is performed for *each* trial value of the other parameters. See Train (1986), pg. 120, for an alternative use of this "calibration" algorithm.

tice). The BLP contraction requires that an independent and identically distributed logit error be appended to consumers' utility functions. Thus, BLP is inconsistent if the addition of this error results in a misspecification of utility. This concern is not just a minor specification quibble, as is demonstrated by the logit welfare numbers from Petrin (2002) and Song (2003), the logit substitution patterns from Goolsbee and Petrin (2004), and the arguments in Hausman and Wise (1976), Berry and Pakes (2001), and Bajari and Benkard (2002).

A third case concerns markets where goods may not all be substitutes. Berry's proof of uniqueness for the product-market controls requires that all of the goods in the choice set are substitutes for one another. Choice sets where some goods may be complements would, for example, be a case where the BLP approach is not available.

For the control function approach, sampling error in market shares is not a primary concern. The approach applies to a wider class of choice sets than those in which goods are all substitutes for one another. It obviates the need for the BLP contraction algorithm, which simplifies computation (for many questions the control function approach can be implemented in standard programming packages). Without the contraction there is also no need to add an i.i.d. logit error, avoiding any specification problems that may arise with it.

6 Three Empirical Applications

Our empirical applications are chosen from markets where industrial organization theory suggests the control function's ability to identify the omitted variables problem and correct it may be most problematic. They span three commonly used data types: aggregate (market-level) data, household-level cross-sectional data, and household-level panel data. For each case we briefly describe the data, the demand side model, the instruments, and different options for control functions. In every case, the control function approach leads one to reject the null that there is no price endogeneity problem. It also provides nearly identical estimates of elasticites to the BLP benchmark (which are very different from the uncorrected approach).

6.1 Automobiles (the BLP case study)

We begin with the original BLP (1995) example: price endogeneity in the automobile market. The application is identical to the reported BLP case in almost every respect: data, demand specification, instruments, and estimation. The only difference is that we do not use a supply side model when we estimate the demand side model (so our point estimates only exactly match their estimates for the cases they examine without the supply side).[24]

6.1.1 Data and Demand Specification

The application uses the same 2217 market-level observations on prices, quantities, and characteristics of automobiles sold in the 20 U.S. automobile markets beginning in 1971 and continuing annually to 1990. The utility function used in BLP is[25]

$$u_{ij} \quad = \quad \alpha ln(y_i - p_{mj}) + \delta_{mj} + \textstyle\sum_k \sigma_k \nu_{ik} x_{mjk} + \epsilon_{ij},$$

where

$$\delta_{mj} = \beta_0' x_{mj} + \xi_{mj}. \tag{11}$$

α is the marginal utility of income parameter and income is assumed to follow a log-normal distribution.[26] Characteristics x_{mj} include a constant term, the ratio of horsepower to weight, interior space (length times width), whether air-conditioning is standard (a proxy for luxury), and miles per dollar. The random coefficients on characteristics are assumed to be normally distributed and independent across characteristics with the mean (β_{0k}) and variance (σ_k) (so ν_{ik} are the mean zero standard normal deviates).[27] ϵ_{ij} is i.i.d. extreme value, so the differenced errors - what's relevant for choice data - are distributed i.i.d. logit.

[24] We focus on the demand side for three reasons: it makes the comparison more transparent, most researchers do not impose a supply side model when estimating demands, and the results are easier to replicate. We could not think of a reason that adding the supply side model would change our finding that both setups produce similar estimates.

[25] Consumer i is in one and only one market m, and $m(i)$ is a function of i (we do not explicitly write m in i's presence below).

[26] The mean varies annually and the variance assumed to be constant across the twenty years.

[27] A variance term is included for the constant to allow for heterogeneity in taste for the outside good.

The control function specification is similar, and is given by

$$u_{ij} = \alpha ln(y_i - p_{mj}) + \beta_0' x_{mj} + \sum_k \sigma_k \nu_{ik} x_{mjk} + f_j(\mu_m; \lambda) + \eta_{mj} + \epsilon_{ij}. \quad (12)$$

The only difference is, without the δ's, $\beta_0' x_{mj}$ is included directly, along with the control function and the new error component.

6.1.2 Instruments

Important for both BLP and the control function approach are the determination of prices in the automobile market. BLP consider an equilibrium pricing function of the general form from (3). They follow the literature and assume that observed product characteristics (except price) are uncorrelated with unobserved characteristics (ξ_{mj}, ω_{mj}). (3) implies that in any market m every product characteristic affects every price in the market, so any product characteristic is a valid instrument for any price. This leads to an abundance of instruments, most of which are likely to be very weak. Pakes (1994) derives the first order basis for the optimal instruments, which amounts to three instruments for each demand characteristic: the characteristic itself (because characteristics are exogenous), the sum of the characteristic across own-firm products (excluding that product), and the sum of the characteristic across rival firm products. The intuition comes from the first order conditions of the oligopoly pricing equilibrium (from BLP, pg. 855):

> products that face good substitutes will tend to have low markups, whereas other products will have high markups and thus high prices relative to cost. Similarly, because Nash markups will respond differently to own and rival products, the optimal instruments will distinguish between the characteristics of products produced by the same multi-product firm versus the characteristics of products produced by rival firms.

With 5 characteristics per vehicle, this yields 15 instruments for each product, and we denote this vector $\tilde{z}_{mj}$.

6.1.3 Price Residuals and the Control Function

The first step of the control function approach is to construct an estimate of the expected price for each product conditional on all exogenous factors observed by the econometrician. With the automobile data, very few observations are available on the same nameplate (i.e. the same product) over time, because cars change characteristics and/or exit. This means some restrictions on $E[p_j \mid z_m, d_m]$ across vehicles will be necessary. Some possibilities include assuming that the expected price function is the same across vehicles in the same year, or across similar vehicles, or both. We make a stronger assumption, imposing that the parameters of this function are the same across all products and all years. This yields 2217 observations on this one function.

A second issue arises because of the abundance of arguments in this function (similar to the abundance of instruments in BLP). We follow the logic outlined in Pakes (1994) and use as arguments for each product j the 15 regressors given by $\tilde{z}_{mj}$, which reflect both demand and cost factors relevant for each product. The only demographic variable is average annual income, and it has little effect on the predicted values for price, so we define

$$\mu_{mj} = p_{mj} - E[p_j \mid \tilde{z}_{mj}],$$

and estimate the expectation using ordinary least squares.[28] Because the expectation is estimated with error, an additional source of error arises (the error in μ_{mj}) that must be accounted for in the standard errors; we describe the correction in Appendix A.

The dimensionality problem also arises with the specification for $f_j(\mu_m; \lambda)$. We consider two parsimonious specifications that are based on the assumption that the own-product residual is principally a function of the own-product unobserved factor. For the first specification, only the own-product residual μ_{mj} from the pricing function enters utility for product j. λ_m, the parameter scaling the price residual, is allowed to vary by year, so $f_j(\mu_m; \lambda) = \lambda_m \mu_{mj}$. We use a normal random deviate with common variance across vehicles and years to approximate $\eta_{mj} = \xi_{mj} - \lambda_m \mu_{mj}$.[29]

[28] A second order approximation yielded nearly identical results.

[29] Note that the specification already has a normal random deviate on a constant term, so this new deviate is not separately identified from the variance of the deviate on the

The second control function specification we use has three terms in the control function and adds only three new parameters. It is given by

$$f_j(\mu_m; \lambda) = \lambda_1 \mu_{mj} + \lambda_2 (\sum_{k \neq j, k \in J(j)} \mu_{mk}) + \lambda_3 (\sum_{k \notin J(j)} \mu_{mk}).$$

The motivation for this control function is similar to the motivation for the instruments in Pakes (1994). The first term is the own-product residual (here with one parameter common across years given as λ_1). The sum of other products' price residuals may also contain information on the magnitude of the own-product's unobserved demand factor (conditional on all observed factors). We use the same two sums that are proposed for pricing instruments; the sum of all of the other residuals of the products made by the same firm, or $(\sum_{k \neq j, k \in J_f} \mu_{mk})$, where J_f is the set of products produced by the firm that produces the product j, and the sum of all the residuals of all the products made by other firms, or $(\sum_{k \notin J_f} \mu_{mk})$. Again, we assume the error component is distributed normal with common mean and variance across years.[30]

6.1.4 Estimation and Results

The estimation approach for BLP starts with candidate values of parameters (α, σ), where σ is the vector of σ_k's. The contraction algorithm locates the $\delta' s$ that match observed to predicted market shares for all 2217 automobiles (the logit error ensures that it converges). These product-market controls are then used in an instrumental variables regression for equation (11) to obtain estimates of β_0. The residuals from the estimated equation (11) are then interacted with the instruments to generate the moments that enter the GMM objective function. Minimizing over (α, σ) is achieved by iterating over these steps.

Estimation for the control function approach proceeds in two steps. In the first step estimates of μ_{mj} are obtained. In the second step the likelihood function is maximized. Common parameters for both of the control

constant term. Thus, the estimated variance on the constant term reflects the sum of the variance on the constant term (heterogeneity in taste for outside good) and the variance of $\xi_{mj} - \lambda_m \mu_{mj}$.

[30] Other specifications (for example) might allow the residuals of cars "close" in product space to j to enter the utility for j. Or one might allow the rescaling parameter and/or the random deviates accounting for η_{mj} to vary by groups of similar cars.

function specifications are $(\alpha, \beta_0, \sigma)$. Specification one includes 20 additional parameters, $(\lambda_{71}, \ldots, \lambda_{90})$, each indexed by the year of the data. The second specification, motivated by Pakes (1994), includes three additional parameters $(\lambda_1, \lambda_2, \lambda_3)$.

The point estimates and standard errors from these specifications are reported Table A1 (in Appendix A), and Table 1 translates these estimates into elasticities. The first column uses the uncorrected logit specification from Column 1 of Table III in BLP (1995); because the data sets are the same, these are the same elasticities that result from the coefficients of their Table III. As they report, ignoring price endogeneity severely biases price elasticities towards zero; overall, 67% of them are inelastic.

Columns 2, 3, and 4 report, respectively, specifications one and two of the control function approach and the BLP approach. Column 2, which uses only the own-product price residual with coefficients that vary by year, is very similar to the Column 3 results, which use the three functions of the price residuals and three coefficients (common across years). Both are very similar in almost every respect to the BLP results in Column 4. Across the corrected specifications no automobile price elasticity is inelastic, and the median elasticity is -2.16 for the BLP case, and either -2.08 or -2.23, depending on which control function specification is examined. The one difference is that the spread of elasticities is slightly larger for BLP, with a one standard deviation spread of 0.19 vs. 0.10 for the control function approaches. All of the results from Columns 2-4 strongly contrast with the uncorrected results from Column 1; for example, at -0.77, the median uncorrected elasticity is only one-third that from the corrected approaches.

BLP report elasticities for selected automobiles from 1990, so we do the same, choosing every fourth automobile from their Table III, in which vehicles are sorted in order of ascending price (the overall average elasticities for 1990 are again very similar between BLP and the control function specifications, and substantially different from the uncorrected approach). The discrepancies between the individual elasticities across the three approaches are small; the absolute value of the difference between BLP and the second control function specification for the Mazda 323, Honda Accord, Acura Legend, and BMW 735i are respectively 0.02, 0.10, 0.05, and 0.03. The discrepancy in the spread of elasticities across all vehicles is also smaller for 1990, as the standard deviations are now 0.14 for the control function

Table 1

Automobile Elasticities: Uncorrected, Control Functions, and BLP

	No Correction[1]	Control Function (1)	Control Function (2)	BLP
Results for 1971-1990				
Median	-0.77	-2.08	-2.23	-2.16
Mean	-1.04	-2.08	-2.22	-2.17
Standard Deviation	0.76	0.10	0.10	0.19
No. of Inelastic Demands	67%	0%	0%	0 %
Elasticities from 1990[2]				
Mean	-1.24	-2.11	-2.24	-2.22
Standard Deviation	0.83	0.14	0.14	0.20
No. of Inelastic Demands	53%	0%	0%	0%
1990 Models (from BLP, Table VI):				
Mazda 323	-0.44	-1.82	-1.94	-1.92
Honda Accord	-0.82	-2.10	-2.27	-2.17
Acura Legend	-1.67	-2.25	-2.37	-2.42
BMW 735i	-3.32	-2.06	-2.21	-2.24

Notes: The uncorrected specification is that from Table III of BLP (1995). 1990 is the year BLP focus on for the individual models; we choose every fourth automobile from their Table VI (the other elasticities were also very similar). The first control function specification allows λ to vary by year; the second specification follows Pakes (1994) (as defined in the text).

approaches vs. 0.2 for the BLP approach. Overall, the corrected approaches in this application yield very similar elasticity estimates and reject the "no correction" results from Column 1.[31]

6.2 Multi-Channel Video (Television)

Our second example applies the uncorrected and both correction methods to *households'* choice among television reception options in 2001, where Goolsbee and Petrin (2004) have emphasized the importance of omitted attributes.

6.2.1 Data and Demand Specification

The specification is similar to Goolsbee and Petrin (2004). Four alternatives are available to households: (1) antenna only, (2) expanded basic service, (3) expanded basic cable with a premium service added, such as HBO, and (4) satellite dish. The data used is a sample of 11,810 households in 172 geographically distinct markets, where each market contains only one cable franchise. The utility specification with product-market controls is given as:

$$U_{ij} = \delta_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} 1_{ig} + \gamma_j' d_i + \sigma \nu_i c_j + \epsilon_{ij}, \qquad (13)$$

where all of the elements of utility that do not vary within a market are subsumed into the product-market controls, which are a function of price and other observed attributes:

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

x_{mj} are the observed characteristics of the product (including a product intercept term) and ξ_{mj} is the unobserved characteristic. The price effect varies across five income groups, with the lowest income group taken as the base and the binary variable 1_{ig} indicating whether household i is in income group g.[32] Demographic variables for household i are given by d_i and enter each choice j with a separate coefficient vector γ_j. A random coefficient is included to allow for correlation in unobserved utility over the

[31] The control function specification nests the uncorrected specification, so one formal test asks whether the λ's from the control function approach enter significantly.

[32] The price coefficient for a household in the lowest income group is α while that for a household in group $g > 1$ is $\alpha + \theta_g$.

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_i is an i.i.d. standard normal deviate. The coefficient σ is the standard deviation of the random coefficient, reflecting the degree of correlation among the non-antenna alternatives, and ϵ_{ij} is i.i.d. extreme value.[33]

For the cable television case, there are 172 markets with four choices in each market, so the dimensionality problem does not arise. Since price does not vary across geographic location for antenna-only and satellite, we are not able to construct price residuals for these products, and we do not include a control function for them (the price term is captured in the antenna and satellite intercept). In the general specification, the control function for expanded-basic has the expanded basic price residual and the premium price residual as arguments, and similarly for the premium control function, so the control function is given as $f_j(\mu_m; \lambda) = \overline{\lambda}_j' \mu_m$. The normal deviate η_{mj} is included to account for the difference in the control function value and the actual value of ξ_{mj}.[34] Otherwise, utility is specified as above, but without the product-market controls, so $\beta_0' x_{mj}$ and αp_{mj} enter directly, as does the control function:

$$U_{ij} = \alpha p_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} 1_{ig} + \beta_0' x_{mj} + \gamma_j' d_i + \sigma \nu_i c_j + \overline{\lambda}_j' \mu_m + \eta_{mj} + \epsilon_{ij}. \quad (14)$$

The Forrester survey provides various demographic characteristics. In the estimation we include family income, household size, education, and type of living accommodations. The survey also includes an identifier for the household's television market, which can be used to link households exactly to their cable franchise provider (whether they subscribe to cable or not).

The cable system information comes from Warren Publishing's 2001 Television and Cable Factbook. The attributes we include, which vary over

[33]The error specification in Goolsbee and Petrin (2004) uses a fully flexible multivariate normal specification in place of the logit error. Here we use a logit error specification (with the random coefficient common across multichannel video alternatives) to stay close to the BLP approach that's applied in practice.

[34]Similar to the automobile case, we do not try to separately identify the variance of this deviate from the variance of normal deviate used for correlation in taste across the three multi-channel video options.

markets, 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 expanded basic service, the price of premium service, and the number of over-the-air channels available. Many of the cable operators are owned by multiple system operators (MSO's) like AT+T and Time-Warner, and we include MSO dummy variables. As mentioned earlier, satellite prices do not vary geographically, and the price of antenna-only is assumed to be zero, so the price variation that is used to estimate elasticities arises from the cable alternatives. For the price of satellite, we use $50 per month plus an annual $100 installation and equipment cost. More details are given in Appendix B.

6.2.2 Instruments

For both approaches we use Hausman (1997)-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 expanded-basic cable and the price of premium cable. 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 demand attributes.

6.2.3 Price Residuals and the Control Function

The data for television choices is different from the automobile data in one important dimension; we observe 172 observations on each type of television product. We can thus estimate the expected price functions product by product. For the base specification we construct the price residuals for expanded basic by regressing the expanded-basic price on all the product attributes listed above for both choices plus both Hausman (1997)-type price instruments (the expanded basic and the premium one). Premium residuals are constructed in a similar manner. Again, since the μ_{mj} are estimated, this stage adds an additional source of error for which we must account.

Many alternatives to the first order approximation for the control function specification are available in this case because the ratio of markets to products is high. We experimented with a number of them. When constructing residuals, we included average demographics in the first stage pricing equations in addition to the product characteristics and instruments. In the likelihood maximization stage we experimented with many different specifications for entering the price residuals. We used a series expansion on the residuals (both signed and unsigned), entered the price residuals with random coefficients on them, and interacted the price residuals with other variables. In every case the extra generality did not result in elasticities that differed much from the control function specification described above. In fact, the price residuals for expanded basic and premium were highly collinear, so we could not reject the more parsimonious specification that just included the own-product residual (the elasticities were virtually identical). We report results for just this simplest base specification given by $f_j(\mu_m; \lambda) = \lambda_j \mu_{mj}$.

6.2.4 Estimation and Results

For the BLP approach we estimate the model with product-market controls using the contraction procedure to solve for the 516 (172*3) additional parameters (conditional on parameters θ not captured in the δ_{mj}'s).[35] The value of the likelihood function is then computed at this value of $(\theta, \delta(\theta))$, and the function is maximized over θ. After the likelihood function is maximized, the $\hat{\delta}_{mj}$'s are regressed on the product attributes using 3SLS. A separate equation is used for the expanded-basic cable and premium cable, with the coefficients of the product attributes constrained across equations (consistent with the usual differentiated products approach). These parameter estimates are reported in Appendix B in Table B2.

Estimation of the control function approach proceeds first by obtaining estimates of the price residuals (as described above). Then, the likelihood function is maximized using the equation for utility from (14).

Table 2 gives price elasticities from the models for each approach. Without any correction for price endogeneity the correlation between price and the unobserved characteristics is so strong that demands are upward sloping (consumers like to pay more). Parameter estimates and standard errors from both the control function approach and the BLP approach reject the uncorrected model. The elasticities from these approaches are very similar,

[35] One control in each market is normalized out.

Table 2

Television Choice Elasticities: Uncorrected, Control Function, and BLP

	No Correction	Control Function	BLP
Price of expanded-basic cable			
Antenna-only share	U	0.96	0.79
Expanded-basic cable share	p	-1.18	-0.97
Premium cable share	w	0.99	0.88
Satellite share	a	0.95	0.87
Price of premium cable	r		
Antenna-only share	d	0.60	0.52
Expanded-basic cable share		0.65	0.57
Premium cable share	S	-2.36	-2.04
Satellite share	l	0.64	0.58
Price of satellite	o		
Antenna-only share	p	0.43	0.42
Expanded-basic cable share	i	0.48	0.43
Premium cable share	n	0.48	0.45
Satellite share	g	-3.79	-3.59

with expanded basic at either -0.97 or -1.18, premium at -2.04 or -2.36, and satellite at -3.59 or -3.79.[36]

6.3 Margarine

Our final example uses *household-level panel* data to estimate the demand for margarine. The framework and data exactly follows that outlined in Chintagunta, Dube, and Goh (2003), who demonstrate that unobserved brand characteristics result in a price endogeneity problem for margarine.[37] They employ product-market fixed effects to control for the unobserved brand characteristics. We apply both the fixed effects and the control function approach here.[38]

6.3.1 Data and Demand Specification

The data are weekly purchase histories of 992 households between January 1993 and March 1995 and were collected by Nielsen for the Denver area using checkout-counter scanners. The data for margarine are restricted to the 16 oz. category and the four observed products are Blue Bonnet, I Can't Believe It's Not Butter (ICBINB), Parkay, and Shedd's. Weekly prices and marketing mix variables - including whether the product is on display and whether it is featured - are recorded for every product available in these categories for all 117 weeks. Posted prices may respond to changes in shelf-space, the availability of in-store coupons, or promotions in complementary or substitute categories, all of which are unobserved by the econometrician. Omitted inventories, if correlated across households because of persistence in prices, can also lead to a correlation in price and the unobserved demand shock.

The utility specification for the fixed effects model is given as:

$$U_{ijt} = \delta_{jt} + \alpha_i p_{jt} + \beta_{ij} + \beta_{iF} F_{jt} + \beta_{iD} D_{jt} + \epsilon_{ijt}. \qquad (15)$$

[36]We also tested for specification issues unrelated to the control function, including random coefficients of other variables, and other types of error components. The simpler specifications for either of the corrected approaches could not be rejected, and the elasticities were virtually unchanged.

[37]They also show this to be true for orange juice.

[38]We are greatly indebted to JP Dube for running the control function specification with his data.

All of the elements of utility that do not vary for product j in week t are subsumed into the fixed effects, so

$$\delta_{jt} = \alpha_0 p_{jt} + \beta_{0j} + \beta_{0F} F_{jt} + \beta_{0D} D_{jt} + \xi_{jt},$$

where p_{jt} is the posted price for brand j at time t, F_{jt} and D_{jt} are indicators that are on if the brand is on feature or display respectively at time t, and ξ_{jt} is the unobserved brand-time characteristic. The common-across-consumers price sensitivity term is given by α_0, and similarly the brand specific intercepts and feature and display intercepts are given by β_{0j} $j = 1, \ldots, 4$, β_{0F}, and β_{0D} respectively. Consumer specific tastes vary around these mean taste parameters and are given by α_i, β_{ij} $j = 1, \ldots, 4$, β_{iF}, and β_{iD}, which are mean zero multivariate normal draws. These random taste coefficients freely vary and covary across price, feature, display, and the brand intercept terms (seven factors), adding a total of 28 additional parameters that summarize the variance covariance matrix of unobserved taste heterogeneity. Finally, ϵ_{ijt} is i.i.d. logit.

For the control function approach, utility is specified as above, but without the fixed effects, so β_{0j} $j = 1, \ldots, 4$ and $\alpha_0 p_{mj}$ enter directly, as does the control function:

$$U_{ijt} = (\alpha_0 + \alpha_i) p_{jt} + (\beta_{0j} + \beta_{ij}) + (\beta_{0F} + \beta_{iF}) F_{jt} + (\beta_{0D} + \beta_{iD}) D_{jt} + \lambda_j \mu_{jt} + \eta_{jt} + \epsilon_{ijt}. \tag{16}$$

Wholesale prices are the instruments for the reported shelf price. The price instruments vary weekly for each brand of margarine. These instruments are appropriate if, for example, the unobserved promotional activities at the retail level are uncorrelated with the wholesale price. For the control function specification, the estimator for μ_{jt} is the residual from the regression of a product's retail price at time t on an intercept, the list price at the wholesale level, and the discount off list price (at wholesale).[39] Since the number of products is small relative to the number of markets, we are again able to enter all of the residuals into each product's expression for utility. However, we found again that elasticity estimates were very similar when only the own-product residual entered, and it was the only residual entering significantly. For the results we report the specification where each product has its own coefficient for its residual, so the full control function is given

[39] Wholesale prices of the other products at time t did not enter significantly.

Table 3
Margarine Own-Price Elasticities: Uncorrected, Control Function, and Fixed Effects

	No Correction	Control Function	Fixed Effects
Blue Bonnet	-1.74	-2.09	-2.05
ICan'tBINB	-4.64	-5.33	-5.44
Parkay	-2.69	-3.31	-3.34
Shedd's	-3.32	-4.23	-4.20

as $f_j(\mu_t; \lambda) = \lambda_j \mu_{jt}$. A separate normal deviate η_{jt} is included to account for the difference in the control function value and the actual value of ξ_{jt} for each product. The residuals are allowed to freely vary and covary across the four products.

6.3.2 Results

Table 3 gives price elasticities across the three models. Appendix C reports the point estimates and standard errors for each of the three specifications. Without any correction for price endogeneity the correlation between price and the unobserved brand characteristics is strong enough for margarine such that the own-price elasticities are substantially underestimated with no correction. Across brands they increase between 20-35% with either correction. In each of these cases the control function and the fixed effects approach provide elasticity estimates that are very similar: -2.09 vs. -2.05, -5.33 vs. -5.44, -3.31 vs -3.34, and -4.23 vs. -4.20.[40] We now turn to Monte Carlo experiments.

7 Monte Carlo Experiments

We construct Monte Carlo data for several situations in which there is an unobserved attribute. The experiments utilize variations on a formulation

[40]Dube reported to us similar findings using orange juice, that is, without a correction, results are biased down, but either correction produces similar elasticities (see early versions of their paper for exact details of their orange juice specification, which is similar in flexibility to the margarine specification described here).

suggested to us by Wolak (2003). A product is sold in each of several markets and its attributes and price vary over the markets. Each consumer lives in one market and either buys or does not buy the product offered in that market. The utility that consumer i who lives in market m obtains from the product is $U_i = \beta_0 + \beta_1 x_m + \beta_2 \xi_m - \alpha p_m + \epsilon_i$, where x_m is a product attribute that is observed by the researcher, ξ_m is a product attribute that is not observed by the researcher, and p_m is the price of the product. The idiosyncratic part of utility, ϵ_i, is assumed to be distributed logit (once normalized to the outside good) and independent of all other factors, so that the share of consumers buying the product in market m is given by the standard logit formula $s_m = \frac{exp(\beta_0+\beta_1 x_m+\beta_2 \xi_m-\alpha p_m)}{1+exp(\beta_0+\beta_1 x_m+\beta_2 \xi_m-\alpha p_m)}$. Price is set at a markup over marginal cost, $p_m = mc_m + markup_m$, with the form of the markup specified in each of the experiments. Two variables affect cost but not demand: w_m, which is observed by the researcher, and ω_m which is not.

We conduct Monte Carlos for five different cases for which the product-market controls are consistent. Parameters for each case are chosen so the model generates a wide range of prices and market shares. We assume the researcher only observes x_m, w_m, p_m and s_m in each market. Data are generated for 1000 markets and the demand parameters are estimated using maximum likelihood for two cases: the standard logit case with no correction, and the control function approach with $\mu_m = p_m - E[p|x_m, w_m]$ (we use OLS to estimate $E[p|x_m, w_m]$). Estimation is repeated 100 times (that is, on 100 different datasets, each with 1000 markets).

As we noted earlier, it is necessary to include an error component to account for the difference between the true unobserved demand factor and its estimated value. While the five Monte Carlos we examine here are motivated by those outlined in Wolak (2003), an important difference is that our specifications *include* an error component, whereas his specifications do not. We find the inclusion of this error component explains some of the differences in what we find and what he reports.

Table 4 summarizes the results. The top part of the table describes the data generated by each of the five cases, and the bottom part of the table reports the average of the parameter estimates and their standard deviation, where all averages are taken across the 100 Monte Carlo runs. The first row includes the average maximum and minimum price and the second row reports the same for market shares. A measure of the skewness of demand generated by any Monte Carlo sample is given by the percentage of 1000 markets where less than 1% of consumers purchase the good. The third row reports the average across Monte Carlo samples of this percentage. The fourth row is the average of the correlations of price and the unobserved product attribute.

The second half of Table 4 reports the parameter estimates for the uncorrected and the control function approach. First, the three parameters for the uncorrected logit model are reported, which illustrates the severity of the inconsistency in every case considered (the true values are $\alpha = 1$, $\beta_0 = 10$, and $\beta_1 = 1$, as described below). Then, the five parameters associated with the control function specification are reported, where the two new parameters are λ, the coefficient on the residual μ_m, and σ, the variance of the error component. We now describe each of these cases in turn.

7.1 Case 1: % markup over cost, product attributes enter marginal cost.

We begin by illustrating a case for which the control function is consistent, and then move to cases where the theory suggests the control function is increasingly likely to perform poorly. The specification is given as

$$\begin{aligned} x_m, \xi_m, w_m, \omega_m \quad & iid \quad N(0,1) \\ U_i &= 10 + x_m + \xi_m - p_m + \epsilon_i \\ mc_m &= 10 + x_m + \xi_m + w_m + \omega_m \\ p_m &= 1.1 * mc_m. \end{aligned}$$

Except for the intercept, all of the parameters of the utility and cost function equal 1. We set the intercept at 10, yielding shares that average close to 0.33, where the curvature of the logit function is reasonably high. Observed and unobserved characteristics are assumed to be independent draws from a normal distribution with a variance of 1. All product attributes, both observed and unobserved, affect marginal cost linearly. The intercept of marginal cost is 10, which insures that marginal cost is positive for all draws of the random variables.

We begin with price as a percentage markup over marginal cost (we explore markups that depend on the demand elasticity next). Price is linear

Table 4
Monte Carlo Results for a Variety of Data Types
Uncorrected (Logit) and Control Function Model Parameter Estimates

Case	1	2	3	4	5 (Wolak)
Data Summary For Monte Carlo Samples					
Price Range (Max/Min)	18.14/3.87	20.73/5.81	33.31/6.77	27.88/11.14	323.65/11.02
Share Range (Max/Min)	0.98/0.003	0.70/0.002	0.72/0.001	0.38/0.001	0.61/0.001
% Share < 0.01	1%	0.6%	2%	6%	24%
$\text{Corr}(p_m, \xi_m)$	0.50	0.55	0.20	0.03	0.005
Parameter Estimate (Standard Deviation)					
	True values: $\alpha = 1$, $\beta_0 = 10$, $\beta_1 = 1$				
Uncorrected (Logit)					
α	0.62	0.44	0.58	0.86	0.63
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
β_0	5.93	3.87	5.47	8.37	5.76
	(0.14)	(0.14)	(0.28)	(0.33)	(0.36)
β_1	0.59	0.44	0.58	0.88	0.95
	(0.03)	(0.02)	(0.04)	(0.05)	(0.08)
Control Function					
α	0.97	0.96	0.94	0.99	0.69
	(0.04)	(0.06)	(0.08)	(0.04)	(0.05)
β_0	9.72	9.58	9.49	9.86	6.24
	(0.41)	(0.64)	(0.99)	(0.47)	(0.56)
β_1	0.96	0.96	0.94	1.04	0.95
	(0.05)	(0.06)	(0.10)	(0.06)	(0.07)
λ	0.44	0.63	0.50	0.08	0.04
	(0.03)	(0.05)	(0.08)	(0.03)	(0.01)
σ	0.22	0.27	0.40	0.21	0.73
	(0.15)	(0.18)	(0.24)	(0.14)	(0.33)

The price ranges and share ranges give the average maximum and minimum prices for 1000 markets across the 100 Monte Carlo runs. % Share < 0.01 is the fraction of markets (out of 1000) that have market shares less than 0.01 (averaged across the 100 datasets). Parameter estimates are the average across the 100 runs, and the standard deviation is reported in parentheses.

in $x_m$and ξ_m, so the resulting distribution of μ_m is joint normal with ξ_m, which satisfies the conditions described in Section 3 for the control function approach to be consistent. Thus, regardless of the choice of parameter values for marginal cost and utility, for the percentage markup, and/or for the variance of the observed and unobserved demand characteristics, the control function approach will be consistent for this case.

The range of the data generated by these parameters is substantial. The average minimum and maximum prices are 3.87 to 18.14, the average minimum and maximum shares are 0.003 to 0.98, and the average share is 0.33. The distribution of market shares is not severely skewed towards zero, as only 1% of markets on average have less than 1% market share. The correlation between the unobserved demand characteristic and price is reasonably high at 0.50.

Maximum likelihood estimation of a logit model with p_m and x_m as the explanatory variables performs very poorly. The average estimate of the price coefficient across 100 runs is 0.62, with a standard deviation of 0.01, and the other parameter estimates are similarly biased. The control function approach yields estimates that are all close to their true values, and all reasonably precisely estimated (the price coefficient is estimated to be 0.97 with a standard deviation of 0.04). Finally, $\hat{\lambda}$ enters with an estimate of 0.44 and a standard deviation of 0.03, rejecting the uncorrected model in favor of a model that has an unobserved demand factor.

7.2 Case 2: Elasticity based markups, all product attributes enter marginal cost.

The specification is the same as in Case 1 except that price is set at $p_m = mc_m + \frac{s_m}{\partial s_m / \partial p_m}$, where the markup is profit-maximizing if the firm takes product attributes as given and faces no potential competition.[41] As table 4 indicates, the range of the data is again substantial. The average market share is 0.24 and shares are not skewed towards zero, and the correlation between price and the unobserved attribute is 0.55. Here the uncorrected approach fares even worse than case 1, with a point estimate on price of 0.44 and a standard deviation of 0.01.

This type of markup creates a complex relation between p_m and ξ_m

[41] The markup is determined iteratively, since s_m and the derivative depend on p_m.

such that the exact distribution of the error component conditional on μ_m is difficult to derive. In this case the control function must be viewed as an approximation. A simple control function approach nevertheless performs well. Again, $\hat{\lambda}$ enters significantly, flagging the specification problem. The average estimate of the price coefficients is equal to 0.96 with a standard deviation 0.063. The other point estimates are similarly close to the truth. The only real difference with case 1 is that the standard deviations have increased by about 50% (approximately). This example shows that elasticity-based markups (that is, markups not linear in characteristics) do not necessarily render a simple control function approach inaccurate, even when there is a high correlation between the unobserved characteristic and price, and when market shares and price vary substantially across markets.

7.3 Case 3: Elasticity-based markup, unobserved demand attribute does not affect marginal cost.

In the previous situations, all observed and unobserved product attributes affect marginal cost. Here the idea is to investigate whether the control function performance deteriorates when ξ_m only enters the markup (and not also the marginal cost), so it has a smaller impact on price. We specify marginal cost as $mc_m = 10 + (x_m + w_m + u_m)$. Cases 1 and 2 can be viewed as examples where there is some correlation between the unobserved demand and cost error, with the unobserved cost error equal to the sum $\xi_m + \omega_m$. Viewed in these terms, this specification rules out correlation between the unobserved demand and cost error, which is likely to work against the control function approach, as discussed earlier. The correlation between the unobserved attribute and price now falls to 0.20. Again, price variation is substantial, as the average maximum price is almost six times the average minimum price, and shares range (on average) from 0.001 to 0.72, with an average mean share of 0.25.

The uncorrected logit model continues to perform poorly, with the point estimate for the price coefficient on average equal to 0.58 (with standard deviation 0.02), and other parameter estimates similarly biased down. The control function approach continues to perform well, again rejecting the uncorrected model in favor of the unobserved attribute alternative. The price coefficient is estimated to be 0.94, with standard deviation 0.11, and the other two coefficient estimates are also close to their true values.

7.4 Case 4: Elasticity-based markup, unobserved demand attribute does not affect marginal cost, marginal costs are exponential.

We now increase the non-linearity in the relationship between price and marginal cost by defining marginal costs as $mc_m = 10 + exp(x_m + w_m + u_m)$. Here we first estimate the model with the variances of the random terms equal to 1/2 (we consider variances equal to 1 in case 5). The generated data has similar price variance, with minimum/maximum averages of 11.15 and 28.07. However, the non-linearity in marginal costs and its affect on price combined with the fact that the cost variables do not affect consumer demand leads to very low demand in some markets, as high prices often do not reflect demand attributes consumers value (prices are driven by observed and unobserved cost factors). On average, 6% of markets have demand shares less than 1%. This same phenomenon leads to a low correlation between price and the unobserved demand attribute, which falls to 0.03 for this Monte Carlo model.

The logit model is still inconsistent in this case, although the price coefficient is less biased relative to earlier cases, partly as a consequence of the fall in the correlation between price and the unobserved attribute (the point estimate is 0.86 and the standard deviation is 0.02). Note that, for this case, it is no longer the correlation between price and the unobserved attribute that is the main cause of the specification problem for the uncorrected approach. Instead, it is the missing unobserved demand attribute combined with the high non-linearity of the system.

The control function approach again identifies the specification problem, as $\hat{\lambda}$ enters significantly. Despite the substantial non-linearities in demand across markets, the control function appears to fully correct the price and other coefficients, as they are all estimated to be quite close to their true values.

7.5 Case 5: Case 4, with increased variance in random terms.

In his base specification Wolak (2003) suggests a variant of this case, which combines all of the features of the previous examples. In particular, the variance is increased back to 1 for all random variables. The data generated from this exercise reflect the non-linearities of the specification: the average maximum price is almost 30 times the minimum price (323 vs. 11), reflecting the widely varying cost shocks. Since these unobserved cost factors do not enter consumers' utility function, on average 240 markets out of 1000 in every sample have market shares of less than 1% (almost 500 in 1000 markets in every sample have market shares less than 5%). Finally, the non-linearities have all but eliminated a correlation between price and the unobserved demand attribute, as it falls to 0.005. As with case 4, the econometric problems arise because there is an unobserved attribute, and not because it is correlated with price.

The control function approach continues to identify the consistency problem, as $\hat{\lambda}$ enters the specification significantly. Both the uncorrected approach and the control function estimates are significantly different from the true parameter values in this case. While we are not aware of a product that exemplifies this kind of price/market share pattern across markets, it does illustrate the potential dangers associated with the control function approach with unobserved product attributes. It also illustrates that even when the control function approach is inconsistent, it can still uncover misspecification in the uncorrected approach.[42]

8 Conclusion

In applications of differentiated product models it is rare that all relevant product attributes are observed by the econometrician. When some attributes are omitted, price will typically be correlated with the unobserved portion of consumers' utility for that product: the equilibrating mechanism in the market causes the price to be higher for products that display desirable attributes observed by consumers and producers but not measured by the econometrician. The positive correlation between price and the unobserved attribute biases estimated price elasticities towards zero.

In this paper we develop a control function approach for omitted at-

[42] Wolak (2003) provides a more striking example that this one. For example, when we replicate his results (using a logit error instead of a probit error for convenience), we find his specification leads to an average maximum and minimum price of 141 and 1 respectively (a ratio of 141).

tributes. In the context of differentiated products, the idea is to add to the demand side model proxies for these unobserved demand characteristics. We exploit the information that prices contain on unobserved demand factors. Our control function has as arguments the differences between product prices and their predicted values given all of the relevant demand and supply factors that the econometrician observes. We show the conditions under which these differences can provide a basis for the unobserved demand factors, so they can be proxied out of the estimation equations.

The control function approach we outline is straightforward to implement in standard programming packages; the first stage is a regression, and the second stage is maximization of a likelihood function. While the approach is not as robust to omitted attributes as including a control for each product in every market, it is consistent for some well-known and pragmatic pricing rules. Even in the cases where the basis for the control function does not fully span the space of unobserved demand attributes, the approach still provides an easy-to-implement test that will often have high power in the case when price endogeneity is a problem, and the estimates are always useful as starting values for any of the more demanding approaches. It is also available in some cases when the product-market controls estimator is not available.

We present three empirical demand applications that replicate specifications from earlier works - all of which use product-market controls - including Berry, Levinsohn, and Pakes (1995), who look at automobiles, Goolsbee and Petrin (2004), who look at cable television, and Chintagunta, Dube, and Goh (2003), who look at margarine. These applications are chosen to span a range of markets and potential competitive pricing behaviors, and include three types of different data: aggregate (market-level) data, household-level cross-sectional data, and household-level panel data. We use them to show in practice how one implements the control function approach for these different types of products and aggregation levels. We test for the presence of price endogeneity by comparing the control function approach to a standard uncorrected approach and reject the null of no price endogeneity in every case. The estimated elasticities are almost identical across the BLP and control function approaches and differ significantly from the uncorrected elasticity estimates, which are severely biased down in every case.

9 Appendix A: Automobile Case-Study Details

Table A1 contains the estimated demand parameters and standard errors for the automobile data. These parameters yield the reported elasticities in Table 1. The first column of estimates is the specification reported in column one of Table III in BLP (1995), where the dependent variable is the log of good j's market share minus the log of the outside good's market share. This log-odds ratio is regressed on price and characteristics to estimate the parameters of the utility function (this specification has no random coefficients). The price parameter is sufficiently biased towards zero to result in 67% of the estimated price elasticities being inelastic, which is inconsistent with profit maximizing behavior.[43] We emphasize again that these parameter estimates and inelastic elasticities have already been reported in BLP (1995) (these results are virtually identical to the results reported there because the data set has almost been perfectly replicated).

The parameter estimates from the control function approach and BLP approach respectively are reported next. The demand specification and data are identical to BLP (1995). The specifications include random coefficients on the characteristics, and price and income enter as $ln(y_i - p_{mj})$. The price parameter for BLP we obtain here is similar to that reported in their second specification in Table IV. Again, the BLP results reported here do not impose the supply side model during estimation, and are thus not identical to their point estimates in Table IV.

For the control function approach, $\hat{\mu}$ are used to approximate μ in the estimation routine, so the standard errors from the traditional formulas (and output by standard estimation routines) are biased downward. To approximate this additional source of variance in the control function approach, we bootstrap the price regressions. Specifically, we reestimate the expected price with a bootstrapped sample, calculate the implied residuals, and re-estimate the model with these new residuals (otherwise using the original data). We then repeat this exercise over many bootstrapped samples. The variance in the parameter estimates across the bootstrapped samples is then added to the variance from the traditional formulas (which are appropriate

[43]These price (and other) parameters are not directly comparable to the parameters from the control function and BLP specifications.

Table A1

Estimated Parameters for Automobile Demand: Uncorrected, Control Function, and BLP

Parameter	Variable	No Correction*	Control Function	BLP
Term on Price (α)	price	-0.088		
		(0.004)		
	$ln(y-p)$		29.743	23.565
			(0.828)	(0.341)
Means ($\overline{\beta}$'s)	Constant	-10.071	-4.319	-6.768
		(0.252)	(0.115)	(27.781)
	HP/Weight	-0.122	1.851	-1.157
		(0.277)	(0.032)	(3.076)
	Air	-0.034	0.548	-0.067
		(0.072)	(0.033)	(2.657)
	MP$	0.265	-0.150	0.260
		(0.043)	(0.004)	(18.624)
	Size	2.342	2.100	3.272
		(0.125)	(0.009)	(37.989)
Std. Deviations (σ_β's)	Constant		0.022	0.003
			(0.005)	(0.322)
	HP/Weight		0.048	3.817
			(0.020)	(0.173)
	Air		0.001	1.233
			(0.069)	(0.059)
	MP$		0.001	0.001
			(0.001)	(6.794)
	Size		0.008	0.033
			(0.002)	(0.081)
Control Function (λ's)	λ_1		0.065	
			(0.003)	
	λ_2		-0.002	
			(0.001)	
	λ_3		0.001	
			(0.001)	

The demand specification and data are identical to BLP (1995). Column 1 is virtually identical to results reported in Table III. We do not impose the supply side model, so column 3 is not identical to their results reported in Table IV, although some coefficients (including price) are very similar.

when μ is observed without error).[44]

[44] Karaca-Mandic and Train (2002) provide a formula for the asymptotic standard errors in this type of two-step estimation.

10 Appendix B: Television Case-Study Details

The information on households' television choices, the characteristics of households, and the prices and attributes of the cable franchise serving the household's geographic area 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. 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.

Table B1 gives the estimated parameters and standard errors. Since $\hat{\mu}$ are used to approximate μ in the estimation routine, the standard errors from the traditional formulas (and output by standard estimation routines) are biased downward. To approximate the additional source of variance arising from using $\hat{\mu}$, we add a new term to the estimated variance of the parameters obtained from treating $\hat{\mu}$ as the true μ. This new component comes from bootstrapping the price regressions. That is, we repeatedly estimate the price regressions with bootstrapped samples, calculate the residuals, and re-estimate the model with the new residuals. The variance in parameter estimates over the bootstrapped price samples is added to the variance estimates from the traditional formulas (which are appropriate when μ is observed without error). These total standard errors are given in the table. The adjustment is important for the standard errors of the base price coefficient, the coefficients for the residuals, and the coefficients of the product attributes, which increase between 50-100%. As noted earlier, Karaca-Mandic and Train (2002) provide a formula for the asymptotic standard errors in this type of two-step estimation; they find that in our application the formula gives standard errors that are very similar to those obtained with the bootstrap procedure.

The first column of Table B1 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 μ_{mj} and the error component are not included. The second column applies the control function approach. 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 unusable 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 income groups, with the magnitude decreasing as income rises.

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. 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 esti-

Table B1: Control Function Approach
to Modeling TV Reception Choice

Alternatives: 1. Antenna only, 2. Basic and expanded cable, 3. Premium cable, 4. Satellite
Variables enter alternatives in parentheses and zero in other alternatives.

Explanatory variable	Uncorrected	With control functions
	(Standard errors in parentheses)	
Price, in dollars per month (1-4)	-.0202 (.0047)	-.0969 (.0400)
Price for income group 2 (1-4)	.0149 (.0024)	.0150 (.0025)
Price for income group 3 (1-4)	.0246 (.0030)	.0247 (.0031)
Price for income group 4 (1-4)	.0269 (.0034)	.0269 (.0035)
Price for income group 5 (1-4)	.0308 (.0036)	.0308 (.0038)
Number of cable channels (2,3)	-.0023 (.0011)	.0026 (.0029)
Number of premium channels (3)	.0375 (.0163)	.0448 (.0233)
Number of over-the-air channels (1)	.0265 (.0090)	.0222 (.0111)
Whether pay per view is offered (2,3)	.4315 (.0666)	.5813 (.1104)
Indicator: ATT is cable company (2)	.1279 (.0946)	-.1949 (.1845)
Indicator: ATT is cable company (3)	.0993 (.1195)	-.2370 (.1944)
Indicator: Adelphia Comm is cable company (2)	.3304 (.1224)	.3425 (.1898)
Indicator: Adelphia Comm is cable company (3)	.2817 (.1511)	.2392 (.2246)
Indicator: Cablevision is cable company (2)	.6923 (.2243)	.1342 (.3677)
Indicator: Cablevision is cable company (3)	1.328 (.2448)	.7350 (.3856)
Indicator: Charter Comm is cable company (2)	.0279 (.1010)	-.0580 (.1441)
Indicator: Charter Comm is cable company (3)	-.0618 (.1310)	-.1757 (.1825)
Indicator: Comcast is cable company (2)	.2325 (.1107)	-.0938 (.2072)
Indicator: Comcast is cable company (3)	.5010 (.1325)	.1656 (.2262)
Indicator: Cox Comm is cable company (2)	.2907 (.1386)	-.0577 (.2496)
Indicator: Cox Comm is cable company (3)	.5258 (.1637)	.0874 (.2954)
Indicator: Time-Warner is cable company (2)	.1393 (.0974)	-.0817 (.1507)
Indicator: Time-Warner cable company (3)	.2294 (.1242)	-.0689 (.1891)
Education level of household (2)	-.0644 (.0220)	-.0619 (.0221)
Education level of household (3)	-.1137 (.0278)	-.1123 (.0280)
Education level of household (4)	-.1965 (.0369)	-.1967 (.0372)
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 (.0886)
Household rents dwelling (4)	-.2129 (.1562)	-.2149 (.1569)
Single family dwelling (4)	.7622 (.1523)	.7649 (.1523)
Residual for expanded-basic cable price (2)		.0805 (.0416)
Residual for premium cable price (4)		.0873 (.0418)
Alternative specific constant (2)	1.119 (.2668)	2.972 (1.057)
Alternative specific constant (3)	.1683 (.3158)	2.903 (1.487)
Alternative specific constant (4)	-.2213 (.4102)	4.218 (2.386)
Error components, standard deviation (2-4)	..5087 (.6789)	.5553 (.8567)
Log likelihood at convergence	-14660.84	-14635.47
Number of observations: 11810		

mates 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), expanded-basic cable, premium cable, and satellite. Larger households tend not to buy expanded-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 generally less able to install a cable hookup or mount a satellite dish. 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 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.

The results for the BLP approach are given in Table B2. The bottom part of the table gives the estimates of the demographic coefficients from the first stage. The top part of the table gives the results of the regression of the product-market controls 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 income 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

Table B2: BLP Approach
to Modeling TV Reception Choice

Alternatives: 1. Antenna only, 2. Basic and expanded 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		

and becomes positive as expected when the endogeneity is corrected. Generally, the coefficients on the product attributes are similar to the control function estimates.

The demographic coefficients in Table B2 are also similar to those from the control function approach. Education induces households to buy less TV reception. Larger households tend not to buy expanded-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.

11 Appendix C: Margarine Case-Study Details

Table C1 contains the estimated demand parameters and standard errors for the margarine data. These parameters yield the reported elasticities in Table 3. In addition to the parameters listed in the table, there are 28 additional parameters that are associated with the fully flexible multivariate normal taste distribution across the seven variables: price, the four brands, and the feature and display variable. Since the variance covariance matrix of η_{mj} is not separately identified from the variance in taste for the brand, we do not estimate a separate variance term for each product (it is absorbed into the brand variance covariance matrix). The fixed effects approach has 117*4=468 additional parameters, or 464 more than the control function specification, which has four additional parameters relative to the uncorrected approach, one for each of the brand residuals: $\lambda_{BB}, \lambda_{IC}, \lambda_{PA}, \lambda_{SH}$.

The first column of estimates is the standard logit model with the additional taste heterogeneity, but without controls for price endogeneity. The next columns report coefficient estimates for the control function and the fixed effects approach respectively. The point estimates associated with price are very similar, at -74.55 and -73.98, and are approximately 25% larger than the price coefficient from the standard logit model.[45] The control function parameters enter significantly and differ across each of the four products, although moving to a common control function parameter (not reported

[45] The standard errors in this model are biased down because there is no correction for the sampling variance arising from the estimated first stage price equation, as we did for the automobile and television cases that we estimated ourselves. Almost all of the first stage regression parameters were precisely estimated, suggesting this source of variance is probably small.

Table C1

Estimated Parameters for Margarine Demand: Uncorrected, Control Function, and Fixed Effects

Parameter	Variable	No Correction	Control Function	Fixed Effects
Term on Price (α_0)	price	-59.88	-74.55	-73.98
		(2.30)	(3.48)	(5.48)
Brand means (β_0's)	Blue Bonnet	-1.90	-1.22	-1.50
		(0.10)	(0.13)	(0.20)
	I Can't BINB	1.11	2.56	2.38
		(0.22)	(0.35)	(0.54)
	Parkay	-0.73	-0.04	-0.29
		(0.13)	(0.18)	(0.29)
	Shedds	-1.04	-0.30	-0.33
		(0.15)	(0.22)	(0.33)
Promotional controls (β_0's)	Feature	0.17	0.20	0.23
		(0.06)	(0.06)	(0.06)
	Display	1.38	1.28	1.59
		(0.27)	(0.29)	(0.30)
Control Function (λ's)	λ_{BB}		17.94	
			(1.74)	
	λ_{IC}		43.71	
			(7.14)	
	λ_{PA}		2.39	
			(4.09)	
	λ_{SH}		10.66	
			(1.77)	
Log-likelihood		-23703	-23633	-23021
Total trips		56138	56138	56138
Total households		992	992	992

Note: All three specifications include a fully-flexible normal variance co-variance matrix for taste heterogeneity across the seven variables (a total of 28 parameters): price, four brands, and feature and display.

below) only changed the price coefficient to -72.36.

References

Bajari, P., and C. Benkard (2002): "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches," Working Paper, Department of Economics, Stanford University.

——— (2003): "Demand Estimation with Hetergeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," Working Paper, Department of Economics, Stanford University.

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.

——— (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.

Berry, S., O. Linton, and A. Pakes (2003): "Limits Theorems for Estimating Parameters of Differentiated Product Demand Systems," *forthcoming, Review of Economic Studies.*

Berry, S., and A. Pakes (2001): "Estimating A Pure Hedonic Discrete Choice Model," Working Paper, Yale University.

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.

Brownstone, D., and K. Train (1999): "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, 89, 109–129.

CARROLL, R., D. RUPPERT, AND L. STEFANSKI (1995): *Measurement Error in Nonlinear Models.* Chapman and Hall.

CHINTAGUNTA, P., J. DUBÉ, AND K. GOH (2003): "Targeted pricing and the estimation of consumer choice models in the presence of unmeasured product characteristics," Working Paper, Graduate School of Business, University of Chicago.

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 (2004): "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica*, 72(2), 351–382.

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.

——— (1997): "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by R. Gordon, and T. Bresnahan. Chicago: University of Chicago Press.

HAUSMAN, J., AND D. WISE (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 48, 403–429.

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 Commuter Travel Time Savings for Car Drivers in New Zealand: Evaluating Alternative Model Specifications," *Transportation*, 28, 101–118.

HONG, H., AND E. TAMER (2003): "A Simple Estimator for Nonlinear Error in Variable Models," *Journal of Econometrics*, 117, 1–19.

KARACA-MANDIC, P., AND K. TRAIN (2002): "Standard Error Correction in Two-Step Estimation with Nested Samples," Working Paper, Department of Economics, University of California, Berkeley.

KOTLER, P. (2002): *Marketing Management*. New Jersey: Prentice Hall.

MCFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press, pp. 105–42.

——— (1978): "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. Amsterdam: North-Holland, pp. 75–96.

——— (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden. Cambridge, MA: MIT Press.

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.

PAKES, A. (1996): "Dynamic structural models, problems and prospects: mixed continuous discrete controls and market interaction," in *Advances in Econometrics, Sixth World Congress, Volume II*, ed. by C. Sims, pp. 171–259, New York. Cambridge.

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

RIVERS, D., AND Q. VUONG (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39, 347–366.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

Shim, E., and E. Sudit (1995): "How Manufacturers Price Products," *Management Accounting*, 76, 37–39.

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.

Song, M. (2003): "Measuring Consumer Welfare in the CPU Market: an application of the pure characteristics demand model," Georgia Institutes of Technology, Working Paper.

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

——— (1998): "Recreation Demand Models with Taste Variation," *Land Economics*, 74, 230–239.

——— (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.

Trajtenberg, M. (1989): "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 97, 444–479.

——— (1990): *Economic Analysis of Product Innovation: The Case of CT Scanners*. Harvard University Press.

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

Wolak, F. (2003): "Review of 'Omitted Product Attributes in Discrete Choice Models'," Presentation at NBER Winter IO meeting.

Yang, S., Y. Chen, and G. Allenby (2003): "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 0, 1–25.