Control Function Corrections for Unobserved Factors in Differentiated Product Models

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

Preliminary

February 15, 2006

Abstract

Unobserved factors in differentiated product models can generate severe bias in price elasticities. We develop a generalized control function method and specification test for this setting based on the non-parametric identification results from Petrin (2005), who shows the assumptions under which price functions can be inverted to obtain controls that condition on the part of the demand error that is not independent of price. Unlike using product-market controls, our approach does not require additive separability between observed and unobserved factors. We develop a "hybrid" approach that also loosens this restriction for that setting. We compare results across approaches on three data sets and demand specifications estimated elsewhere that span a range of markets and levels of aggregation, including automobiles (the original Berry, Levinsohn, and Pakes (1995) application), cable television, and margarine. The estimated elasticities are similar across the control function and product-market control approaches, and they both differ significantly from the uncorrected elasticity estimates, which are significantly biased up in every case.

1 Introduction

Models of differentiated products are widely used for estimating demand elasticities and substitution patterns. In applications of these models it is rare that all relevant factors are observed by the econometrician. When some factors are unobserved, price will typically be correlated with these unobserved factors through the equilibrating mechanism in the market. For example, products that display desirable attributes observed by consumers and producers but not measured by the econometrician will often have prices that are positively correlated with the demand error. Alternatively, if advertising or other promotional activities are omitted from the specification, and if prices are set simultaneously with these promotional levels, then price will be correlated with the demand error.[1] The problem has arisen in both aggregate (i.e. market-level) data and disaggregate (i.e., customer-level) data, and empirically has tended to bias estimates of price elasticities in a positive direction.

Since demand in discrete choice settings is not linear in price, standard linear methods for correcting this endogeneity problem are not immediately applicable. In this paper we develop a generalized control function approach for endogenous prices in differentiated-product discrete-choice models. Unlike the product-market control solutions of Berry (1994) and Berry, Levinsohn, and Pakes (1995), we do not require additive separability between observed and unobserved factors, so advertising campaigns or other promotional activities (e.g.) can affect the marginal impact of price or a product characteristic on utility. We then extend our method to a hybrid "control-function/product-market control" that also does not impose additive separability.

We exploit the information that prices contain on unobserved factors. The specifications are based on results from Petrin (2005), who extends Imbens and Newey (2003) to demand settings, showing the conditions under which demand is non-parametrically identified using control functions. The

[1] In this case the bias could be positive or negative, depending upon how prices are set with promotional activities.

intuition underlying the correction in non-linear settings is similar to that in a linear environment, where a new control variable is included in the regression to condition out the part of the error that is correlated with the endogenous regressor (see the ideas discussed in Telser (1964), which are more formally developed in Heckman (1976), Heckman (1978), and Hausman (1978)).[2] The approach we adopt inverts prices to recover random variables that are one-to-one functions with the error in the reduced form pricing equation. Using these variables as controls in the demand equation conditions out the dependence of price on the demand error. Finally, we develop a χ^2 portmanteau specification test that has power against the assumptions maintained by the control function approach.

We are not the first to look to the information that prices contain on demand unobservables (see e.g. Trajtenberg (1989,1990), Villas Boas and Winer (1999), Bajari and Benkard (2005), and Petrin and Train (2005).) Our approach is perhaps most closely related to Bajari and Benkard (2005). They extend the differentiated products setting of Rosen (1974) to an imperfectly competitive one, providing restrictions on the demand and supply side such that price is *only* a function of own-product observed and unobserved demand characteristics, and price is monotonic in the unobserved factor. Finally, if "many products are observed in a single market," they can estimate the pricing equation to recover a one-to-one function of the unobserved demand factor; they compare the prices of products with identical observed demand characteristics and attribute differences in prices to differences in the unobserved demand factor.

Our control function approach generalizes their setup to differentiated-product settings where the equilibrium pricing function cannot be expressed solely as a function of a product's own observed and unobserved characteristics, and where product space is not full of "many" products. Specifically, unlike the Bajari and Benkard (2005) setting, we can allow the supply side to affect equilibrium prices, and we can allow consumers to have idiosyncratic tastes for products, so the widely used generalized extreme value (e.g.

[2]It has been applied to a Tobit model by Smith and Blundell (1986) and binary probit by Rivers and Vuong (1988).

logit and nested logit) and (multivariate) normal models are not ruled out.

Our control function approach has its roots in a broad literature on nonparametric estimation of demand (see e.g. Brown (1983), Roehrig (1988), Brown and Matzkin (1998), Benkard and Berry (2004), and Matzkin (2005)), and relates closely to the literature using restrictions from economic theory for identification in semi- and nonparametric settings, as in Matzkin (1994), Olley and Pakes (1996), Levinsohn and Petrin (2003), and Matzkin (2003), and to the literature on identification in nonseparable models (see Blundell and Powell (2001) and Chesher (2003)).

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 have been chosen because uncorrected price elasticities have been shown to suffer from severe bias, and because they 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.

Our approach is to choose simple specifications for demand in our three applications that are comparable across the product-market control approach and the control function approach in order to see whether parsimonious corrections yield similar results that are significantly different from performing no correction. We leave for future work exploration of more complicated specifications that explore when differences arise between the approaches. Our simplest formulations for the control function approach use regression in the first stage and then maximize a likelihood in the second stage, and can thus be run in standard programming packages. For simple demand specifications the elasticities are almost identical across the the control function approach and the Berry, Levinsohn, and Pakes (1995) approach, and they differ significantly from the uncorrected elasticity estimates, which are biased down in every case.

The paper proceeds as follows. Section 2 sets up a structural demand model, describes in this setting how bias arises from unobserved factors, and then develops the control function approach, the product-market approach, and the hybrid model. Section 3 establishes conditions under which the reduced form pricing function exists and is invertible, and discusses estimation of the controls. Section 4 reports the results from the three applications using real data, and Section 5 concludes.

2 The Demand Model and Approximations

We assume there are $k = 1, \ldots, J$ goods, and consumers purchase good j if $u_j > u_k \,\forall k \neq j$. The choice probability for good j is given as

$$s_j = \int \{u_j \geq u_k \,\forall k \neq j\}\, f_u(u \,|\, X, P)\, du,$$

where $u = \langle u_1, \ldots, u_J \rangle$, and $f_u(u \,|\, X, P)$ denotes the density of u conditional on the entire vector of observed product characteristics $X = (X_1, \ldots, X_j)$ and prices $P = (P_1, \ldots, P_J)$ (we abstract from observed individual characteristics here).[3] We let capital letters denote random variables and lower-case letters denote realizations. The important question from the approximation standpoint is how one should write down a flexible parametric functional form for $f_u(u \,|\, X, P)$ that does not impose strong separability between observed and unobserved product-specific factors, is consistent in the presence of endogenous prices, and allows for idiosyncratic consumer-specific tastes for product characteristics.

To simplify exposition we assume there is only one observed product characteristic per product (in addition to price). We also maintain the standard assumption that consumers only derive utility from the characteristics of the good that they actually consume. Utility consumer i derives from good j is then given as

$$u_{ij} = u(x_j, p_j, \varepsilon_{ij}).$$

[3]This is without loss of generality as adding a vector s of individual characteristics to the specification simply entails rewriting the density as $f_u(u \,|\, X, P, s)$ with s as one of the conditioning vectors.

The error may reflect unobserved product and consumer specific factors that affect utility, including unobserved physical characteristics, advertising and promotional activities of which the researcher is unaware, and idiosyncratic consumer-specific tastes for observed and unobserved factors.

As McFadden (1981) argues, there is little loss in generality of assuming a utility structure that is linear in variables and parameters, because the variables can be potentially complex transformations of the arguments entering utility, and one can approximate any continuous indirect utility function on a compact set arbitrarily well using a linear-in-parameters specification. We write u_{ij} as linear and second order flexible in observed and unobserved factors:

$$u_{ij} = \beta_0 \, p_j + \beta_1 \, x_j + \beta_2 \, p_j \, x_j + \beta_3 p_j^2 + \beta_4 x_j^2 + \beta_5 p_j \varepsilon_{ij} + \beta_6 x_j \varepsilon_{ij} + \varepsilon_{ij}, \quad (1)$$

We include the interactions and squared terms because, without them, underlying preferences would have to be strongly separable in the arguments of the utility function (see Goldman and Uzawa (1964)), requiring marginal rates of substitution between observed and unobserved factors to be independent of the levels of consumption of each factor. For example, unobserved advertising campaigns or other promotional activities may change willingness to pay, here given as the marginal utility of price, and this will likely impact the marginal rate of substitution between x and p, violating strong separability. The specification is easily extended to allow for more flexibility along these dimensions.

To complete the utility specification we must characterize the distribution of $f_{\varepsilon | X,P}(\varepsilon \,|\, X, P)$. We assume there exists a product-specific vector of random variables, denoted $\xi = (\xi_1, \ldots, \xi_J)$, such that after conditioning on ξ, the density of ε is mean independent of (X, P), as in Berry, Levinsohn, and Pakes (1995) (heretofore BLP). Specifically, we reformulate (1) as

$$\begin{aligned} u_{ij} &= \beta_0 \, p_j + \beta_1 \, x_j + \beta_2 \, p_j \, x_j + \beta_3 p_j^2 + \beta_4 x_j^2 + \beta_5 p_j \xi_j + \beta_6 x_j \xi_j + \xi_j \\ &+ \sigma_p p_j \epsilon_{ip} + \sigma_x x_j \epsilon_{ix} + \epsilon_{ij}. \end{aligned} \quad (2)$$

with parameters $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$ and $\sigma = (\sigma_x, \sigma_p)$. The utility specification maintains the second order flexibility of (1), but restricts it to observed and unobserved product-specific factors.[4] The remaining variance is then ascribed to a $J+2$ element vector

$$\epsilon_i = (\epsilon_{ip}, \epsilon_{ix}, \epsilon_{i1}, \ldots, \epsilon_{iJ}),$$

which has one idiosyncratic taste draw for each observed product characteristic and one idiosyncratic taste draw for each product. This vector is assumed to be a mean zero vector of errors that is independent of ξ and of all observed factors. It is also assumed to follow some parametric formulation (e.g. i.i.d. logit, multivariate normal, or some mix of the two). The important difference between this specification and the one proposed in BLP is that this specification does not impose $(\beta_5, \beta_6) = 0$, and thus does not insist on additive separability between observed and unobserved product-specific factors.

2.1 Bias from Unobserved Factors

An econometric problem often occurs because standard choice models maintain that ε is independent of (X, P), e.g. the logit model maintains that the unobserved component of utility is independent of the observed variables:

$$f_{\varepsilon|X,P}(\varepsilon \,|\, X, P) = f_\varepsilon(\varepsilon). \tag{3}$$

GEV models, 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 (for example). If ε in part reflects unobserved factors that are not measured by the econometrician, and if sellers charge prices based on these unobserved factors, then ε and P will not be independent, and parameter estimates under the maintained assumption in (3) will not be consistent.[5]

[4] Using random coefficients it is straightforward to allow different consumers have different second-order expansions.

[5] One example would be if consumers willingness to pay increases in the unobserved factor, and producers charge more for products with more of the unobserved factor, then the price elasticity bias is likely to go in the positive direction; consumers look less price

Since Trajtenberg (1989)'s finding of upward sloping demand curves for CT scanners, numerous empirical applications have shown that ε and P can be so highly positively 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)), yogurt and ketchup (Villas-Boas and Winer (1999)), and margarine and orange juice (Chintagunta, Dubé, and Goh (2003)), to name a few.

2.2 Control Function Approximations to Demand

We return to the formulation of demand from (1), but for expository purposes we focus on a single good "buy or not buy" setup

$$u_i = \beta_0 p + \beta_1 x + \beta_2 p x + \beta_3 p^2 + \beta_4 x^2 + \beta_5 p \varepsilon_{ip} + \beta_6 x \varepsilon_{ix} + \varepsilon_i. \tag{4}$$

Extension to the multiple-good choice set is straightforward and is done throughout the rest of the paper. The remaining specification question then relates to $f_{\varepsilon|X,P}(\varepsilon \mid X, P)$.

The control function appeals to economic theory to determine other equations in which ξ enters as an argument. If ξ is identified from these equations, then one can condition on it to address the endogeneity problem.[6] More generally, if a one-to-one function of ξ is identified - defined as $\tilde{\xi}$ - it is sufficient to condition upon this variable to address the endogeneity problem.

The control function approach proceeds in two stages. In the first stage one recovers $\tilde{\xi}$ for each product by inverting the relevant equation. In the process, one also recovers the complete marginal distribution $f_{\tilde{\xi}}(\tilde{\xi})$. Using the law of iterated expectations we can write $f_{\varepsilon|X,P}(\varepsilon \mid X, P)$ as

$$f_{\varepsilon|X,P}(\varepsilon \mid X, P) \quad = \quad f_{\varepsilon|X,P}(\varepsilon \mid X, P, \tilde{\xi}) \, f_{\tilde{\xi}}(\tilde{\xi}). \tag{5}$$

sensitive than they actually are because they are getting "more" for paying the observed price they pay than the econometrician has taken into account.

[6] In this paper we focus on the reduced form pricing equation, which we discuss at length in Section 3.

From (5), the remaining question is then how to parametrically approximate $f_{\varepsilon|X,P}(\varepsilon \,|\, X, P, \tilde{\xi})$, where the mean of ε will generally depend upon $\tilde{\xi}$ but is now independent of (X, P).

While many alternative approximations are available, we adopt a simple and flexible formulation for

$$\beta_5 p \varepsilon_{ip} + \beta_6 x \varepsilon_{ix} + \varepsilon_i. \tag{6}$$

We let $\tilde{\xi}$ interact with both x and p, and allow for $\tilde{\xi}$ and $\tilde{\xi}^2$ as separate arguments:

$$\lambda_1 p\tilde{\xi} + \lambda_2 x\tilde{\xi} + \lambda_3\tilde{\xi} + \lambda_4\tilde{\xi}^2, \tag{7}$$

with new parameters $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$.[7] We adopt the same formulation for the (mean-zero) variance term from (1), assuming it is given as $\sigma_p p \epsilon_{ip} + \sigma_x x \epsilon_{ix} + \epsilon_i$. The two terms together provide the approximation for (6), and together with (4), the final specification for utility is then given by

$$\begin{aligned} u_i &= \beta_0\, p + \beta_1\, x + \beta_2\, p\, x + \beta_3 p^2 + \beta_4 x^2 + \lambda_1 p\tilde{\xi} + \lambda_2 x\tilde{\xi} + \lambda_3\tilde{\xi} + \lambda_4\tilde{\xi}^2 \\ &+ \sigma_p p \epsilon_{ip} + \sigma_x x \epsilon_{ix} + \epsilon_i. \end{aligned} \tag{8}$$

As long as the parametric approximation of ε in $\tilde{\xi}$ is flexible enough, the approach is consistent in the presence of price endogeneity, it allows for heterogeneity in tastes, and it does not impose strong separability. Intuitively, the approach is consistent because conditioning on $\tilde{\xi}$ holds constant the variation in ε that is not independent of price. The remaining variation in price - which is independent of ε - is then used to learn how varying price affects demand.

If we take the approximation from (1) as the true model, a comparison of (8) with (1) shows that the only difference between the models arises because the product-specific error

$$\lambda_1 p\tilde{\xi} + \lambda_2 x\tilde{\xi} + \lambda_3\tilde{\xi} + \lambda_4\tilde{\xi}^2 \tag{9}$$

is used to approximate the product-specific error

$$\beta_5 p\xi + \beta_6 x\xi + \xi.$$

[7] We could easily allow x and p to interact with $\tilde{\xi}^2$ also.

If (1) is the correct specification, in the limit observed and predicted shares will match exactly. A direct test of specification for the control function model is then to evaluate whether the predicted shares from the estimated model for (8) are statistically different from the observed shares in the data. Indeed, the motivation for including ξ as a new error, as noted by BLP, is that models without product-specific errors often "overfit;" only sampling error can explain the difference between the data and the model, and there is often insufficient variance to account for the discrepancy, leading χ^2 tests for fit to reject the specification. Thus, if the control function specification passes this test and the uncorrected model does not, the main symptom has been addressed by the addition of (7) to the specification. If the test rejects, then either a more flexible formulation for (7) is required, or there is a problem with the control variable $\tilde{\xi}$. Thus, as long as there is reasonable power under the alternative, the test is informative, and its ease of use suggests it should generally be reported if this approach is taken.

2.3 Product-Market Control Approximations to Demand

If utility is given by the specification in (1), the product-market control approach is not identified. To see this, denote the product-market control ξ, and note in the single good "buy or not buy" specification of (1) we have

$$\begin{aligned} u_i &= \beta_0\, p + \beta_1\, x + \beta_2\, p\, x + \beta_3 p^2 + \beta_4 x^2 + \beta_5 p\xi + \beta_6 x\xi + \xi \\ &+ \sigma_p p\epsilon_{ip} + \sigma_x x\epsilon_{ix} + \epsilon_{ij}. \end{aligned} \tag{10}$$

Direct estimation is not possible because there are an infinite number of values of (β, σ, ξ) consistent with any observed data set. Thus, for this specification, the product-market control approach is not very useful empirically.

To shed light on the necessary restriction for identification, it is useful to rewrite (11) as Berry (1994) does, decomposing it additively into three components:

$$u_i = \delta + \mu_i + \epsilon_i, \tag{11}$$

one that is common across consumers ($\delta = \delta(x, p, \xi; \beta)$), one that allows for idiosyncratic consumer tastes for observed product characteristics ($\mu_i =$

$\mu_(x, p, \epsilon_{ix}, \epsilon_{ip}; \sigma))$, and one that allows for idiosyncratic consumer-product specific tastes (ϵ_i). When expressed in this manner, it is clear that (δ, σ) are identified. Given that δ is identified, and given as

$$\delta = \beta_0\, p + \beta_1\, x + \beta_2\, p\, x + \beta_3 p^2 + \beta_4 x^2 + \beta_5 p\xi + \beta_6 x\xi + \xi \qquad (12)$$

it is now clear that insisting on strong separability between observed and unobserved characteristics is sufficient for identification. Intuitively, if we assume $(\beta_5, \beta_6) = 0$, then

$$\delta = \beta_0\, p + \beta_1\, x + \beta_2\, p\, x + \beta_3 p^2 + \beta_4 x^2 + \xi, \qquad (13)$$

and with ξ now additively separable, $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ are identified. Thus, additive separability of the error from observed factors is a necessary condition for identification of any product-market control approach.

2.4 A Hybrid Control-Function/Product-Market Control Approach

As noted in the previous section, the product-market control approach does identify

$$\delta = \beta_0\, p + \beta_1\, x + \beta_2\, p\, x + \beta_3 p^2 + \beta_4 x^2 + \beta_5 p\xi + \beta_6 x\xi + \xi. \qquad (14)$$

If $\tilde{\xi}$ is indeed one-to-one with ξ, then projecting δ onto $(p, x, p\, x, p^2, x^2)$ and a sufficiently flexible specification in $(p, x, \tilde{\xi})$ will produce consistent estimates of $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$. Consistent estimates of price elasticities (e.g.) for any value of $\tilde{\xi}$ can be constructed, as can consistent estimates of the aggregate price elasticity which averages over the marginal distribution of $\tilde{\xi}$.

2.5 Discussion

There are some important estimation and econometric issues that arise with respect to all three approaches that we now explore. First, the latter two approaches include controls for each product in each market. These are effectively fixed effects, meaning the usual fixed-effect concerns are relevant.

Specifically, if the number of observations on the fixed effects is not increasing at a sufficiently rapid rate, the incidental parameters problem (Neyman and Scott (1948)) causes the estimator for the fixed effects to be inconsistent. In the non-linear setting this is particularly problematic, as the bias in the fixed effects is transmitted to all of the parameters being estimated. Additionally, the objective function is neither linear in parameters nor generally globally concave in them, leading to a curse of dimensionality as there are often several hundred extra parameters to estimate with product-market controls. This curse of dimensionality can also be a problem for the control function setting if the specification for (7) is flexible.

Berry (1994) 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 which is applicable for either of the latter two approaches. While useful in many settings, this approach tends to exacerbate the "fixed-effects-in-a-non-linear-setting" problem, because the product market controls are concentrated out by matching observed to predicted shares, leading sampling or simulation error in these shares to enter the estimating equations in a highly non-linear manner. Berry, Linton, and Pakes (2003) work out the exact cost; the estimator is consistent and asymptotically normal if the number of observed purchasers (or simulation draws) increases at an exponential rate relative to the number of products.[8]

There are two other potential advantages to avoiding the contraction algorithm that locates the product-market controls. First, for it to be a contraction an independent and identically distributed (i.i.d.) logit error must be appended to the utility function.[9] This assumption is problematic in many settings, as noted in Petrin (2002), Berry and Pakes (2003), Goolsbee and Petrin (2004), and Song (2003). Second, the computational burden of the approach can be severe, especially when used in settings with micro-

[8]If the number of products is indexed by J, then the number of purchasers and simulation draws must grow at the rate J^2.

[9]The algorithm is likely to work well with an i.i.d. error that does not depart significantly from the logit distribution. The algorithm becomes problematic when the errors begin to covary significantly across the choices, as potentially with a multivariate normal specification (see Goolsbee and Petrin (2004)).

level data and/or random coefficients, where integration up to the product-market shares must be done by simulation.[10] This computational burden has effectively precluded its use by many practitioners and policy-makers.[11]

3 Inverting Prices to Recover Control Variates

We write the reduced form for prices as

$$P = p(Z_1, Z_2, \xi),$$

where $Z = (Z_1, Z_2)$, with $p(\cdot)$ determined by demand shifters Z_1, cost-side factors Z_2, and the error(s) ξ. For now we assume that it is sufficient to consider only one error per product in the reduced form. We establish sufficient conditions for the existence of a reduced form price equation that is invertible in ξ for the single-product monopolist. We then show sufficient conditions for the differentiated product setting with multi-product vendors, where there is a system of J price equations and the vector of product errors ξ has J elements. Finally, we provide estimators for $\tilde{\xi}$, the one-to-one function of ξ. The discussion draws heavily on Petrin (2005), and readers are referred to that paper for more details and results (see also Appendix D).

[10] Because error in the simulated integral enters the estimation equations non-linearly, the simulated integrals must be evaluated for many more simulation draws than is typically done (e.g. thousands). Furthermore, for any candidate value of parameters, evaluation of the objective function requires the calculation of these simulated integrals possibly hundreds of times because the error remaining from the contraction procedure must be sufficiently small, again because this error enters the estimator non-linearly. For alternative approaches that do not use the contraction (like the control function approach) only one evaluation of this integral is necessary, with possibly many fewer simulation draws.

[11] In our view this burden is why the BLP approach has yet to be widely adopted by regulatory agencies, who are often under time-constraints in antitrust/merger cases. On computational burden, all of the heretofore cited papers that use product-market controls are coded in either Matlab or Gauss, and many of them call C routines to speed up the computational algorithm that locates the fixed effects to reduce run-times to a manageable level.

3.1 Single-Product Setting

We start with the single-product monopolist. We broadly define product characteristics to include physical or perceived physical characteristic of the good (like style), and advertising or promotional activities undertaken by firms to encourage demand for the good. Let x be this vector of product characteristics, where the scalar ξ might be one of the characteristics observed to consumers and producers but unobserved to the practitioner. Let z be other potential demand and cost shifters (the remaining elements of Z), $q(p, x, z)$ be quantity demanded at price p given (x, z), and $mc(x, z)$ be the marginal cost of production. Profits are given as

$$\Pi = (p - mc(x, z)) * q(p, x, z), \tag{15}$$

and the optimal (static) price p^* solves

$$\frac{\partial \Pi(p^*, x, z)}{\partial p} = 0. \tag{16}$$

Lemma 1 provides a set of sufficient conditions that relates changes in characteristics to changes in price.

Lemma 1 *Assume* $\frac{\partial q}{\partial p} < 0$, $\frac{\partial mc}{\partial x_k} \geq 0$, *and* $\frac{\partial q}{\partial x_k} \geq 0$. *For product characteristic* x_k, *if*

$$\frac{\partial q}{\partial x_k} - \frac{\partial mc}{\partial x_k} * \frac{\partial q}{\partial p} \quad > \quad -(p^* - mc) * \frac{\partial^2 q}{\partial x_k \partial p}, \tag{17}$$

then $\frac{\partial p}{\partial x_k} > 0$.

Proof

Total differentiation of the first-order condition with respect to p and x_k coupled with the implicit function theorem leads to the comparative static:

$$\begin{aligned} \frac{\partial p}{\partial x_k} &= -\left(\frac{\partial^2 \Pi(p^*)}{\partial p\, \partial p'}\right)^{-1} \frac{\partial^2 \Pi(p^*)}{\partial p\, \partial x_k} \\ &= -\left(\frac{\partial^2 \Pi(p^*)}{\partial p\, \partial p'}\right)^{-1} * \left(\frac{\partial q}{\partial x_k} + (p^* - mc) * \frac{\partial^2 q}{\partial x_k \partial p} - \frac{\partial mc}{\partial x_k} * \frac{\partial q}{\partial p}\right). \end{aligned} \tag{18}$$

The result follows because $-\frac{\partial^2 \Pi(p^*)}{\partial p\, \partial p'}$ is positive from profit maximization. ‖

The first three assumptions are reasonably weak, requiring respectively that demand is downward sloping, marginal cost is increasing in the amount of the characteristic, and demand is increasing in the amount of the characteristic (consumer's value the characteristic). The left hand side of (17) is positive under these assumptions, which means the result depends on $\frac{\partial^2 q}{\partial x_k \partial p}$, which characterizes how the price elasticity of demand changes as the characteristic increases. If demand becomes less price elastic, then the result holds. If demand becomes more price elastic as the characteristic increases, then the left hand side of the inequality must exceed the right hand side. When these sufficient conditions hold, price is monotonically increasing in x_k, which implies both that price can be written as a *function* of the unobserved factor, and that price is invertible in this factor.

3.2 Multi-Product Setting

We focus on static Bertrand-Nash price competition, which is perhaps the most popular maintained competitive assumption in the applied discrete choice literature. Let $\xi = (\xi_1, \ldots, \xi_J)$ denote the vector of unobserved factors, one for each product. For the Bertrand-Nash setup, define $\Pi_f(p, Z, \xi)$ as the profits for firm f, $f = 1, \ldots, F$, which produces a subset of the goods $j \in J_f$ and chooses prices to maximize static profits. Let p enter as the first J arguments and ξ enter as the last J arguments in every profit function. For fixed Z, let $p^* = (p_1^*, \ldots, p_J^*)$ denote a vector that satisfies

$$\frac{\partial \Pi_{f_j}(p^*, Z, \xi)}{\partial p_j} = 0, \quad j = 1, \ldots, J. \tag{19}$$

Lemma 2 *Given Z, if each of the first order conditions in (19) is continuously differentiable in its first and last J arguments, then p^* can be expressed as an implicit function of ξ,*

$$p^*(Z, \xi) = (p_1^*(Z, \xi), \ldots, p_J^*(Z, \xi))', \tag{20}$$

in a neighborhood of ξ. If

$$\frac{\partial^2 \Pi(p^*(Z, \xi), Z, \xi)}{\partial p \, \partial \xi'} \tag{21}$$

has full rank, then the matrix of derivatives $\frac{\partial p}{\partial \xi'}$ is invertible in this neighborhood.

Proof
Total differentiation of the first order conditions from (19) yields

$$\frac{\partial^2 \Pi(p^*(Z,\xi),Z,\xi)}{\partial p\, \partial p'} dp = -\frac{\partial^2 \Pi(p^*(Z,\xi),Z,\xi)}{\partial p\, \partial \xi'} d\xi. \tag{22}$$

The matrix

$$\frac{\partial^2 \Pi(p^*(Z,\xi),Z,\xi)}{\partial p\, \partial p'} \tag{23}$$

is full rank by profit maximization, so the first claim follows directly from the implicit function theorem. Solving for the JXJ matrix of derivatives $\frac{\partial p}{\partial \xi'}$ yields

$$\frac{\partial p}{\partial \xi'} = -\left(\frac{\partial^2 \Pi(p^*(Z,\xi),Z,\xi)}{\partial p\, \partial p'}\right)^{-1} \frac{\partial^2 \Pi(p^*(Z,\xi),Z,\xi)}{\partial p\, \partial \xi'}. \tag{24}$$

$\frac{\partial p}{\partial \xi'}$ is equal to the product of two full rank matrices, which is also full rank and thus invertible. ||

Lemma 2 shows that the conditions under which prices can be written as a one-to-one function of the vector of unobserved factors. Full rank of (23) comes directly from profit maximization. Thus, the invertibility turns on the full rank of (21). This requires the vector of first order conditions to vary in J independent directions when differentiated with respect to the vector ξ.

3.3 Estimation of the Control Variates

Given invertibility, which we now maintain throughout, the final step that remains is to show the conditions under which $\tilde{\xi}$ is identified. In our empirical approaches we will be using repeated observations on markets to recover the control variables, so we add a market index m to the pricing equation:

$$P_{mj} = p_j(Z_m, \xi_m), \tag{25}$$

where $Z_{mj} = (Z_{mj1}, Z_{mj2})$ is the set of the observed characteristics for product j that affect demand Z_{mj1} and costs Z_{mj2}, and all observed characteristics in market m relevant for the determination of prices in equilibrium are included in the vector $Z_m = (Z_{m1}, Z_{m2}, \ldots, Z_{mJ})$.[12] The product-specific errors are denoted $\xi_m = (\xi_{m1}, \ldots, \xi_{mJ})$. m can serve as an index for the same market observed repeatedly over time, a cross-section of markets at a given point in time, or a combination of the two.

In the univariate case, identification of $\tilde{\xi}$ is achieved by extending the estimator proposed in Imbens and Newey (2003) to demand systems. Theorem 1 in Appendix D provides the details, showing that if $p(Z_m, \xi_m)$ is monotonic in ξ, and ξ is independent of Z, then the conditional distribution of P given Z is equal to the marginal distribution of ξ:

$$F_{P|Z}(P|Z) = F_\xi(\xi).$$

In words, if two markets are observationally equivalent and one market has higher prices than the other market, the higher-price market has a higher unoberved ξ. The proof is constructive, suggesting the empirical cumulative distribution function for $F_{P|Z}(P|Z)$ as an estimator for $F_\xi(\cdot)$. Specifically, the control function is defined as $\tilde{\xi} = F_{P|Z}(P|Z)$, a random variable that is one-to-one with ξ and uniformly distributed over the unit interval.

In the multivariate setting, which we focus on in the empirical results, the results for identification from Matzkin (2005) are relevant. She spells out the general conditions for identification that can be used in conjunction with a "Closest Empirical Distribution" estimator (e.g. Manski (1983) and Brown and Matzkin (1998)) to recover the one-to-one function of ξ. Petrin (2005) provides a complete discussion of available options, and Theorems 3 and 4 provide estimators for particular formulations of the reduced form pricing equations (see Appendix D). We exclusively use Theorem 3 in the empirical results later, but emphasize the availability of many alterative approaches.

[12] In a setting with multi-product sellers Z_{mj} will include indicators for sellers of each product because the same set of products divided up differently across the set of producers will generally result in different equilibrium prices.

When there are multiple sellers in the market(s), multiplicity of equilibria can present an econometric problem. If we can condition on the different equilibria, and the functions are allowed to differ across the equilibrium index, then there is no econometric problem. However, if markets where different equilibria are being played are pooled together when estimating the equilibrium pricing function, estimates of the control variates will not be consistent, because the underlying relationship is not a function but instead a correspondence. This problem is likely to be most pronounced when the same market over time is not observed, but instead cross-sections of markets are used in the estimation. This problem is well-known in the literature that estimates dynamic games, where it is usually assumed that, conditional on the covariates, only one equilibrium is played in the data.[13] In terms of testing, if results change when pricing functions are estimated by geographic region (say), or some other observable by which equilibria might differ, pricing functions should be allowed to vary across these observable factors. If control variates are not consistently estimated because of multiplicity of equilibria, the specification test proposed in section 2.2 has power against "no multiplicity", as observed and predicted shares should diverge.

A computationally simple approach to obtaining control variates is to assume that prices are additively separable in the unobserved factors. Because it is so straightforward to implement, we use it in our base comparisons between the control function and product-market control approaches later. Under additive separability, P_{mj} can be written as

$$\begin{aligned} P_{mj} &= g_{1j}(Z_m) + g_{2j}(\xi_m) \\ &= E[\,P_j \mid Z_m] + \mu_j(\xi_m), \end{aligned} \tag{26}$$

where in the second line we define the difference between price and its expected value conditional on observed exogenous factors as $\mu_{mj} = \mu_j(\xi_m)$. Repeated use of OLS can be used to estimate the vector of conditioning variables, given as

$$\tilde{\xi}_m = (\mu_1(\xi_m), \mu_2(\xi_m), \ldots, \mu_J(\xi_m)),$$

[13]See, for example, Aguirregabiria and Mira (2003), Pakes, Ostrovsky, and Berry(2003), Pesendorfer, and Schmidt-Dengler (2003), Seim (2003).

which is one-to-one with ξ_m as Theorem 3 shows.

If in the data the number of observed markets is large relative to the dimension of Z_m, variation in prices and observed characteristics are typically sufficient to identify and estimate the proxies. 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 dimension of Z_m is large relative to the number of observed markets, there may be "too many" regressors to use in the estimation of the residuals. We suggest using an approximation from Pakes (1994), who 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 arguments enter the function without changing the value of the function). 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.

In summary, the control function approach can fail if: we need to allow for more than one error per product, there are multiplicity of equilibria across markets that are not appropriately indexed, and/or the inverse $\tilde{\xi}$ it not identified. We note that the χ^2 test we proposed in section 2.2 has power against any of these potential problems, and thus provides a very useful portmaneau specification test.

4 Three Empirical Applications

Our empirical applications 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. Our approach is to choose simple specifications for demand in our three applications that are comparable across the product-market control approach and the control function approach in order to see whether parsimonious corrections yield similar results that are significantly different from performing no correction. We leave for future work exploration of more complicated specifications that explore when significant differences arise between the approaches.

4.1 Multi-Channel Video (Television)

Our first application 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.

4.1.1 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. Utility is specified as:

$$u_{ij} = \alpha p_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} 1_{ig} + \beta_0' x_{mj} + \gamma_j' d_i + \sigma \nu_i c_j + \lambda_j' \tilde{\xi}_m + \epsilon_{ij}. \qquad (27)$$

x_{mj} are the observed characteristics of the product (including a product intercept term). 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.[14] $\beta_0' x_{mj}$ denotes the base utility derived from observed product characteristics. 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 three non-antenna alternatives. In particular, $c_j = 1$ if j is one of the three non-antenna alternatives and $c_j = 0$ otherwise,

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

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. Finally, $\varepsilon_{ij} = \lambda_j' \tilde{\xi}_m + \epsilon_{ij}$, with λ_j the vector of parameters associated with the controls, and ϵ_{ij} is i.i.d. extreme value.[15]

The utility specification with product-market controls is given as:

$$u_{ij} = \delta_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} 1_{ig} + \gamma_j' d_i + \sigma \nu_i c_j + \epsilon_{ij}, \tag{28}$$

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:

$$\delta_{mj} = \alpha p_{mj} + \beta x_{mj} + \overline{\xi}_{mj},$$

with $\overline{\xi}_{mj}$ the unobserved factor, which we write with the overscore because it is chosen such that

$$s(\theta, \delta(\theta)) = s^{Data},$$

where θ includes all parameters but the product-market control, and δ_{mj} is obtained from matching observed to predicted shares, as discussed in section 2.5.

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

[15]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.

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 A.

4.1.2 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.

4.1.3 The Control Functions

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. Since price does not vary across geographic location for antenna-only and satellite, we are not able to construct price residuals for these products.

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 $\varepsilon_i = \lambda_j \tilde{\xi}_{mj} + \epsilon_{ij}$.

4.1.4 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).[16] 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 A in Table A2.

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 (27).

Table 1 gives price elasticities from the models for each approach. Without any correction for price endogeneity the correlation between price and

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

Table 1

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

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, 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.[17]

4.2 Margarine

Our second application 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 use product-market controls to demonstrate that unobserved brand characteristics result in a price endogeneity problem for margarine.[18]

4.2.1 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

[17]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.

[18]They also show this to be true for orange juice. We are greatly indebted to JP Dube for running the control function specification with his data.

shock.

The utility specification for the control function approach is given as

$$u_{ijt} = (\alpha_0 + \alpha_i)p_{jt} + (\beta_{0j} + \beta_{ij}) + (\beta_{0F} + \beta_{iF})F_{jt} + (\beta_{0D} + \beta_{iD})D_{jt} + \lambda_j' \tilde{\xi}_t + \epsilon_{ijt}. \tag{29}$$

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 $\tilde{\xi}_t$ is the vector of control variates. 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.

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

$$u_{ijt} = \delta_{jt} + \alpha_i p_{jt} + \beta_{ij} + \beta_{iF}F_{jt} + \beta_{iD}D_{jt} + \epsilon_{ijt}. \tag{30}$$

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

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

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 $\tilde{\xi}_t$ is the (vector of) residuals from the regression of each 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).[19] Since the number of products is small relative to the number of markets, we

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

Table 2
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

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 $\varepsilon_i = \lambda_j \tilde{\xi}_{mjt} + \epsilon_{ij}$.

4.2.2 Results

Table 2 gives price elasticities across the three models. Appendix B 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.[20]

[20]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).

4.3 Automobiles (the BLP case study)

Our third example is 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).[21]

4.3.1 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[22]

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

where

$$\delta_{mj} = \beta_0' x_{mj} + \overline{\xi}_{mj}, \tag{31}$$

with $\overline{\xi}_{mj}$ the unobserved factor, which we write with the overscore because it is chosen such that

$$s(\theta, \delta(\theta)) = s^{Data},$$

where θ includes all parameters but the product-market control, and δ_{mj} is obtained from matching observed to predicted shares, as discussed in section 2.5. α is the marginal utility of income parameter and income is assumed to follow a log-normal distribution.[23] Characteristics x_{mj} include

[21] 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.

[22] 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).

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

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).[24] ϵ_{ij} is i.i.d. extreme value, so the differenced errors - what's relevant for choice data - are distributed i.i.d. logit.

The control function specification is similar, and is given by

$$u_{ij} = \alpha ln(y_i - p_{mj}) + \beta_0' x_{mj} + \sum_k \sigma_k \nu_{ik} x_{mjk} + \lambda_j' \tilde{\xi}_m + \epsilon_{ij}. \qquad (32)$$

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.

4.3.2 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 (25). They follow the literature and assume that observed product characteristics (except price) are uncorrelated with unobserved characteristics $\overline{\xi}_{mj}$. (25) 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

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

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 $\tilde{z}_{mj}$.

4.3.3 The Control Functions

We 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 $E[\,P_j \mid Z_m]$ 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) - described above - 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

$$\mu_{mj}(\xi_m) = P_{mj} - E[\,P_j \mid \tilde{Z}_{mj}],$$

and estimate the expectation using ordinary least squares.[25] Because the expectation is estimated with error, an additional source of error arises. We describe the correction for the standard errors in Appendix A.

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

The dimensionality problem also arises with the specification for $f(\varepsilon_i \mid \tilde{\xi}_m)$. 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

$$\varepsilon_i = \lambda_m \mu_{mj} + \epsilon_{ij},$$

where ϵ_{ij} is i.i.d. logit.

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

$$\varepsilon_i = \lambda_1 \mu_{mj}(\xi_m) + \lambda_2 (\sum_{k \neq j, k \in J(j)} \mu_{mk}(\xi_m)) + \lambda_3 (\sum_{k \notin J(j)} \mu_{mk}(\xi_m)) + \epsilon_{ij}.$$

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 $(\sum_{k \neq j, k \in J_f} \mu_{mk}(\xi_m))$, 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 $(\sum_{k \notin J_f} \mu_{mk}(\xi_m))$.[26]

4.3.4 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 (31) to

[26] Other specifications (for example) might allow the residuals of cars "close" in product space to j to enter the utility for j.

obtain estimates of β_0. The residuals from the estimated equation (31) 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 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 C1 (in Appendix A), and table 3 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

Table 3

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).

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 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.[27]

5 Conclusion

In applications of differentiated product models it is rare that all the relevant factors are observed by the econometrician. When some demand factors are omitted, price will typically be correlated with these unobserved factors through the equilibrating mechanism in the market, and this correlation will bias estimated price elasticities.

In this paper we develop generalized control function method and specification test for this setting that does not impose the additive separability between observed and unobserved factors required by the popular product-market control approaches. Our specifications are based on the results from Petrin (2005), who extends Imbens and Newey (2003) to demand settings, showing the conditions under which demand is nonparametrically identified when the errors are not additive. The approach inverts prices to recover a random variable that is a one-to-one function of the unobserved product attribute, and then uses this control to condition out the dependence of the demand error on price. We develop a hybrid "control-function/product-

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

market control" approach that loosens the additive separability assumption for product-market controls, so (e.g.) advertising campaigns or other promotional activities can affect the marginal impact of price or a product characteristic on utility.

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. The estimated elasticities are almost identical across the BLP and control function approaches and differ significantly from the uncorrected elasticity estimates, which are biased upward in every case.

6 Appendix A: 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 A1 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 A1 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

Table A1: 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		

this attribute, unlike the others, is not on a per-channel basis; its coefficient represents the value of the option to purchase pay-per-view events. The point estimates imply that households are willing to pay \$6.00 to \$8.88 per month for this option, depending on their income.

Several demographic variables enter the model. Their estimated coefficients are fairly similar in the corrected and uncorrected models. The estimates suggest that households with higher education tend to purchase less TV reception: the education coefficients are progressively more highly negative for antenna-only (which is zero by normalization), 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 A2. 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 vari-

Table A2: 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		

ables 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 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.

7 Appendix B: Margarine Case-Study Details

Table B1 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.[28] The control function parameters enter significantly and differ across each of the four products, although moving to a common control function parameter (not reported

[28] 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 B1

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 covariance 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.

8 Appendix C: Automobile Case-Study Details

Table C1 contains the estimated demand parameters and standard errors for the automobile data. These parameters yield the reported elasticities in Table 3. 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.[29] 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 reestimate the model with these new residuals (otherwise using the original data). We then repeat this exercise over many bootstrapped samples. The

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

Table C1

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.

variance in the parameter estimates across the bootstrapped samples is then added to the variance from the traditional formulas (which are appropriate when μ is observed without error).[30]

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

9 Appendix D: Selected Results from Petrin (2005)

9.1 Single-Product Markets

Assumption 3.1 *With probability one $p(Z, \xi)$ is monotonic in ξ.*

Assumption 3.2 *(ξ, ε) are jointly independent of Z.*

Theorem 1 is the identification result for the control function.

Theorem 1 *If 3.1 and 3.2 hold, then $F_{P|Z}(P|Z) = F_{\xi}(\xi)$.*

Proof
Let $p^{-1}(z,p)$ denote the inverse of $P = p(z, \xi)$ in its second argument, which exists by assumption A1. Then

$$\begin{aligned} F_{P|Z}(p|z) &= Pr(P \leq p \,|\, Z = z) \\ &= Pr(p(z, \xi) \leq p \,|\, Z = z) \\ &= Pr(\xi \leq p^{-1}(z,p) \,|\, Z = z) \\ &= Pr(\xi \leq p^{-1}(z,p)) \\ &= F_{\xi}(p^{-1}(z,p)), \end{aligned} \tag{33}$$

where the third equality follows from A1 and the fourth line follows from A2. The conclusion follows from A2, which implies $\xi = p^{-1}(Z, P))$. $\|$
The proof is constructive, suggesting the empirical cumulative distribution function for $F_{P|Z}(P|Z)$ as an estimator for $F_{\xi}(\cdot)$. Specifically, the control function is defined as $\tilde{\xi} = F_{P|Z}(P|Z)$, a one-to-one function of ξ. Given this definition, $\tilde{\xi}$ is a random variable that is uniformly distributed over the unit interval.

9.2 Multi-Product Markets

The full triangular system with J goods is given as

$$\begin{aligned} &Q_j = q_j(P, Z_1, \varepsilon), \quad j = 1, \ldots, J \\ &P_j = p_j(Z, \xi) \quad j = 1, \ldots, J, \end{aligned} \tag{34}$$

where $P = (P_1, \ldots, P_J)$ and $\xi = (\xi_1, \ldots, \xi_J)$ denote the vector of prices and errors, Z_1 continues to denote consumer demographics and now potentially all observed product characteristics for the J goods, and the scope of Z and ε are similarly extended. The setup allows, for example, for differentiated products' settings, where observed and unobserved characteristics can in general affect the demand and prices of all of the goods in the market.

Assumption 4.1 *With probability one the JXJ matrix of derivatives $\frac{\partial p}{\partial \xi'}$ is invertible.*

Assumption 4.2 *(ξ,ε) are jointly independent of Z, and ξ_j and ξ_k are independent for $k \neq j$.*

Assumption 4.3 *Prices can be written as additively separable in observed and unobserved factors:*

$$P_j \quad = \quad g_{1j}(Z) + g_{2j}(\xi) \quad j = 1, \ldots, J. \tag{35}$$

Theorem 2 *Let*

$$\tilde{\xi}_j \equiv g_{2j}(\xi) - E[g_{2j}(\xi)] = P_j - E[\,P_j\,|\,Z\,] \quad j = 1, \ldots, J.$$

If 4.1, 4.2, and 4.3 hold, then

$$\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_J)'$$

is a one-to-one function of ξ.

Proof

Invertibility of $P(Z, \xi)$ in ξ (holding Z constant) implies $g_{2\cdot}(\xi)$ is a one-to-one function of ξ, where

$$g_{2\cdot}(\xi) \quad = \quad P - g_{1\cdot}(Z) \tag{36}$$

(all of these objects are $JX1$ vectors). Let the $JX1$ vector of constants denoted K equal

$$E[g_2(\xi)] \quad = \quad (E[g_{21}(\xi)], \ldots, E[g_{2J}(\xi)])'. \tag{37}$$

Then

$$\begin{aligned}\tilde{\xi} &\equiv P - E[P \mid Z] \\ &= P - g_{1\cdot}(Z) - E[g_2(\xi)] \\ &= g_{2\cdot}(\xi) - E[g_2(\xi)]\end{aligned} \tag{38}$$

is also invertible in ξ, because subtracting off a vector of constants from a function invertible in ξ yields a function that is still invertibile in ξ. ‖

We now consider the second case.

Assumption 4.1 guarantees that with probability one $p_j(Z,\xi)$ is monotonic in ξ_j holding Z and ξ_{-j} constant, $j = 1,\dots,J$. Assumption 4.4 requires the unobserved factors to enter the pricing functions in a triangular manner, so one price is a function only of its own unobserved factor, a second price is a function of the first product's unobserved factor plus its own unobserved factor, and so forth.

Assumption 4.4 *Prices satisfy a triangularity assumption in unobserved factors, so for some indexing of products*

$$\begin{aligned}P_1 &= p_1(Z,\xi_1), \\ P_2 &= p_2(Z,\xi_1,\xi_2),\dots \\ P_J &= p_J(Z,\xi_1,\xi_2,\dots,\xi_J).\end{aligned} \tag{39}$$

Theorem 3 *Let $\tilde{\xi}_j = F_{\xi_j}(\xi_j)\ \forall j$, and let $\tilde{\xi}_j^*$ denote the realized value of $\tilde{\xi}_j$. Define $\tilde{\xi}^{j-1} = (\tilde{\xi}_1,\tilde{\xi}_2,\dots,\tilde{\xi}_{j-1})$ and let $\tilde{\xi}^{j-1,*}$ denote the vector of realized values. If 4.1, 4.2, and 4.4 hold, then*

$$\begin{aligned}F_{\xi_1}(\xi_1) &= F_{P_1|Z}(p_1|z), \\ F_{\xi_2}(\xi_2) &= F_{P_2|Z,\tilde{\xi}_1}(p_2|z,\tilde{\xi}_1^*),\dots \\ F_{\xi_J}(\xi_J) &= F_{P_J|Z,\tilde{\xi}_1,\dots,\tilde{\xi}_{J-1}}(p_J|z,\tilde{\xi}_1^*,\dots,\tilde{\xi}_{J-1}^*).\end{aligned} \tag{40}$$

Proof

Let $p_j^{-1}(z,p_j,\xi^{j-1})$ denote the inverse of $P_j = p_j(z,\xi^j)$ in ξ_j. which exists by assumption 4.1b. $F_{\xi_1}(\xi_1) = F_{P_1|Z}(p_1|z)$ follows directly from Theorem

1. The proof proceeds by induction, showing that given $\tilde{\xi}^{j-1}$, $F_{\xi_j}(\xi_j)$ is identified.

$$\begin{aligned} F_{P_j|Z,\tilde{\xi}^{j-1}}(p_j|z,\tilde{\xi}^{j-1,*}) &= Pr(P_j \leq p_j \,|\, Z = z, \tilde{\xi}^{j-1} = \tilde{\xi}^{j-1,*}) \\ &= Pr(p_j(\xi_1,\ldots,\xi_j,Z) \leq p_j \,|\, Z = z, \tilde{\xi}^{j-1} = \tilde{\xi}^{j-1,*}) \\ &= Pr(\xi_j \leq p_j^{-1}(z,p_j,\xi^{j-1}) \,|\, Z = z, \tilde{\xi}^{j-1} = \tilde{\xi}^{j-1,*}) \\ &= Pr(\xi_j \leq p_j^{-1}(z,p_j,\xi^{j-1})) \\ &= F_{\xi_j}(p_j^{-1}(z,p_j,\xi^{j-1})), \end{aligned} \tag{41}$$

where the third equality follows from monotonicity (4.1b), and the fourth line follows from independence of Z and ξ, the independence of ξ_j and ξ_k, and the fact that $\tilde{\xi}^{j-1} = \tilde{\xi}^{j-1,*}$, which makes ξ^{j-1} a deterministic function, so the integral over $f(\xi^{j-1})$ reduces to the integral over $f(\xi_j)$. $\|$
Again, the proof is constructive, suggesting the empirical cumulative distribution function $F_{P_j|Z,\tilde{\xi}^{j-1}}(p_j|z,\tilde{\xi}^{j-1,*})$ as an estimator for $F_{\xi_j}(\cdot)$.

References

AGUIRREGABIRIA, V., AND P. MIRA (2002): "Sequential Simulation Based Estimation of Dynamic Discrete Games," Boston University Working Paper.

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," forthcoming, Journal of Political Economy.

BENKARD, C. L., AND S. BERRY (2004): "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988)," Cowles Foundation Discussion Paper No. 1482.

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.

Blundell, R., and J. Powell (2001): "Endogeneity in Nonparametric and Semiparametric Regression Models," Working Paper, Department of Economics, University College London.

Brown, B. W. (1983): "The Identification Problem in Systems of Nonlinear in the Variables," *Econometrica*, 51(1), 175–196.

Brown, D., and R. Matzkin (1998): "Estimation of Nonparametric Functions in Simultaneous Equations, with an Application to Consumer Demand," Cowles Foundation Working Paper.

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

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71(5).

Chintagunta, P., J. Dubé, and K. Goh (2005): "Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models," *Forthcoming, Marketing Science.*

Crawford, G. (2000): "The Impact of the 1992 Cable Act on Household Demand and Welfare," *RAND Journal of Economics*, 31, 422–449.

Goldman, S., and H. Uzawa (1964): "A Note on Separability in Demand Analysis," *Econometrica*, 32(3), 387–398.

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

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. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475–492.

——— (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.

IMBENS, G., AND W. NEWEY (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper.

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.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 342–371.

MANSKI, C. (1983): "Closest Empirical Distribution Estimation," *Econometrica*, 51(2), 305–320.

MATZKIN, R. (1994): *Restrictions of Economic Theory in Nonparametric Methods*Elsevier Science, chap. 42.

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

——— (2005): "Conditions for Identification in Nonparametric Systems of Equations," Northwestern University Working Paper.

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.

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

Neyman, J., and E. Scott (1948): "Consistent Estimation from Partially Consistent Observations," *Econometrica*, 16.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Industry," *Econometrica*, 64, 1263–1297.

Pakes, A. (1994): "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.

Pesendorfer, M., and P. Schmidt-Dengler (2003): "Identification and Estimation of a Dynamic Game," Working Paper, LSE.

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

——— (2005): "The Use of Control Functions to Identify Demand when Errors are Non-Additive," *University of Chicago GSB Working Paper.*

PETRIN, A., AND K. TRAIN (2005): "Tests for Omitted Attributes in Differentiated Product Models," *University of Chicago GSB Working Paper.*

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

ROEHRIG, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447.

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

SEIM, K. (2002): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," Stanford University Working Paper.

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.

TELSER, L. G. (1964): "Iterative Estimation of a Set of Linear Regression Equations," *Journal of the American Statistical Association*, 59, 845–862.

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.