Tests for Omitted Attributes in Differentiated Product Models

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

April 27, 2005

Abstract

We develop overidentification tests for unobserved demand attributes, exploiting the covariance between the attribute and price when constructing the tests. The tests reject "exogenous prices" if the demand residuals covary with the difference between actual prices and those predicted using observed demand and cost factors. The tests are easy to implement in standard regression packages, as we demonstrate for three recent demand applications with endogenous prices that span aggregate (market-level) data, household-level cross-sectional data, and household-level panel data, and for several Monte Carlos. Our simplest test specifications identify the price endogeneity problem in every case and provide a close estimate of the magnitude of the bias from omitting the attribute.

JEL Classification: C3, L0

Corresponding author: Amil Petrin, Department of Economics, University of Minnesota, Heller Hall 1035, Minneapolis, MN 55455, email: petrin@umn.edu.

1 Introduction

Characteristic-based models of demand are widely used for estimating elasticities and substitution patterns in differentiated product markets. When relevant product attributes are not observed by the practitioner, price can be correlated with the unobserved portion of consumers' utility for that product: producers charge more and consumers are willing pay more for products with more of the omitted attribute, holding all else constant. The positive correlation between price and the unobserved portion of utility biases estimates of price elasticities towards zero. This problem arises for both aggregate (i.e. market-level) data and disaggregate (i.e., customer-level) data, and has been documented empirically for CAT scanners (Trajtenberg (1989)), automobiles (Berry, Levinsohn, and Pakes (1995) and Petrin (2002)), cable television choices (Goolsbee and Petrin (2004) and Crawford (2000)), cereals (Nevo (2001)), yogurt and ketchup (Villas-Boas and Winer (1999)), and margarine and orange juice (Chintagunta, Dube, and Goh (2005)), to name just a few examples.

In this paper we develop a class of overidentification tests that are easy to implement in standard regression packages and have reasonable power against the "exogenous prices" hypothesis. Our approach tests for an unobserved attribute by constructing a proxy for it that exploits the attribute's correlation with price, which is the source of the econometric problem. We develop both conditional moment tests of specification, following Newey (1985), Tauchen (1985), and McFadden (1987), and control function type tests in the spirit of Smith and Blundell (1986), Rivers and Vuong (1988), and Villas-Boas and Winer (1999).[1]

Two alternative approaches are available for testing. One can estimate

[1]The linear control function case is described in Heckman (1978) and Hausman (1978). The first use of the term "control function" of which we are aware is in Heckman and Robb (1985) in the context of selection models.

a variant of the control function models suggested above. This typically requires estimation of the parameters of a bi- or multi-variate normal distribution.[2] A second alternative is to estimate a more general model that protects against the potential endogeneity problem but adds computational difficulty, like the non-linear fixed effects model in Chintagunta, Dube, and Goh (2004) (CDG) or the inversion approach from Berry (1994) and Berry, Levinsohn and Pakes (1995) (BLP).

A major advantage of our testing approaches is their simplicity. All of the tests use Ordinary Least Squares in an initial stage to obtain the proxies, which are calculated as the difference between product price and its predicted value given the relevant demand and supply factors that the econometrician observes. The control function approach includes the proxies directly in the maximization of the likelihood function and tests for their significance. The conditional moment test estimates the likelihood model without the proxies and rejects exogeneity if the proxies covary significantly with the (generalized) residuals, conditional on the estimated scores.[3]

While rejection of "exogenous prices" using either the conditional moment or control function test provides the strongest evidence of endogeneity, one can also view large changes in the point estimates between the specifications with and without the proxy as evidence of a problem, especially if the price coefficient changes substantially. If no evidence of an omitted attribute is found using the overidentification tests, one can avoid the difficulties associated with estimating the more general models. If evidence is found, the resulting estimates from the control function test specification are suggestive of the direction and magnitude of the bias arising from the price

[2]The models above have only been applied to the binary probit/tobit case in the former two papers, and to a choice model with only three choices in the latter case.

[3]Estimated scores and generalized residuals are standard likelihood function estimation output.

endogeneity, although one must be cautious of this interpretation given the non-linear environment.[4] Finally, these point estimates and standard errors can be used to construct ranges for starting values for any more general and more protective approach.

We describe approaches to implementing the tests for three of the recent demand exercises mentioned above, where prices are shown to be endogenous. Each of these exercises uses a different kind of data, including aggregate (market-level) data (BLP), household-level cross-sectional data (Goolsbee and Petrin (2004)), and household-level panel data (CDG). Our empirical application applies the test to the cable and satellite dish demand specification from Goolsbee and Petrin (2004). In separate work (Petrin and Train (2004)) we replicate the automobile demand specification from BLP and the margarine demand specification from CDG. Our Monte Carlo results build in correlation between price and the unobserved attribute by setting prices according to the standard inverse-elasticity rule, where demand depends in part on the unobserved attribute. Our simplest test specifications identify the price endogeneity problem in every case and provide a close estimate of the magnitude of the bias from omitting the attribute.

The paper proceeds as follows. Sections 2-3 describe differentiated products demand models and the endogeneity problem. Sections 4-8 describe the theory of the tests and their implementation. Sections 9-10 include the empirical application and the Monte Carlos.

2 Demand and Omitted Attributes

The problem of omitted attributes arises naturally within characteristics-based demand approximations. At the core of these approaches is a desire

[4]The conditional moment framework can also be used to calculate this bias, although it is more involved than the control function approach, which reports the bias directly.

for parsimony; an unrestricted constant-elasticity-of-demand system with J goods can have J^2 or more parameters (for example). Characteristics' based approaches achieve parsimony in two ways. They assume that demands for J goods can be approximated by $K << J$ characteristics of the goods, where the K factors serve as the basis for utility. They also assume that consumers only derive utility from the characteristics of the actual good that they purchase. A specification problem arises when the K factors used by the econometrician exclude a relevant characteristic.

We use a random coefficients setup to model demand. Utility that consumer i derives from product j is given as

$$U_{ij} = \sum_k x_{jk}\beta_{ik} - \alpha_i p_j + \beta_{i\xi}\xi_j + \epsilon_{ij}, \qquad (1)$$

with x_{jk} and p_{jk} characteristics and price, coefficients varying across individual observations according to (e.g.)

$$\beta_{ik} = \beta_{k0} + \sum_{l=1}^{L} \beta_{kl} d_{il},$$

and l indexing i's vector of characteristics d_i. $\epsilon_i = (\epsilon_1, \epsilon_2, \ldots, \epsilon_J)'$ is independent across individuals $i = 1, 2, \ldots, N$ with joint density denoted $f_\epsilon(\cdot)$.[5] ξ_j is a product-specific characteristic known to both consumers and producers in the market. We isolate it from the other characteristics because the specification question asks whether ξ_j enters utility for product j (i.e. whether $\beta_{i\xi} \neq 0$).

The log-likelihood function is given as

$$logL_N(Y_1, Y_2, \ldots, Y_N | X, p, \xi, d, \theta) = \sum_{i=1}^{N} \sum_j Y_{ij} logP_{ij}(\theta), \qquad (2)$$

[5] d_i can include demographics or other observed characteristics of i. It can also include a vector of unobserved tastes for the characteristics.

where (X, p) is the $J \times K + 1$ matrix of the entire set of characteristics with arbitrary row (X_j, p_j), $\xi = (\xi_1, \ldots, \xi_J)'$, Y_{ij} is an indicator variable that equals one if i is observed to choose j:

$$Y_{ij} = 1\{U_{ij} > U_{ik} \; \forall k \neq j | \theta, X, p, \xi, d_i, \epsilon_i\}, \tag{3}$$

$Y_i = (Y_{i1}, \ldots, Y_{iJ})'$, and $P_{ij}(\theta)$ is the probability i chooses j conditional on d_i

$$\begin{aligned} P_{ij}(\theta) &= \int_\nu Y_{ij} f_\epsilon(\nu) d\nu \\ &= Prob(U_{ij} > U_{ik} \; \forall k \neq j | \theta, X, p, \xi, d_i). \end{aligned} \tag{4}$$

If ξ is excluded from the specification but does enter utility, then the parameter estimates obtained from maximizing the likelihood function are inconsistent even when ξ is independent of X and p because the system of demand equations given by (4) is not linear in ξ.

3 Endogenous Prices

The omitted attribute problem is exacerbated when this attribute is correlated with price. We develop the mechanism for this correlation here and make use of it later when we derive the omitted attribute tests.

In a multi-product environment where firms sell only a single differentiated product the optimal static pricing rule is given as:

$$p_j = mc_j - \frac{q_j}{\partial q_j / \partial p_j}, \tag{5}$$

where mc_j is the marginal cost of producing j.

Lemma 1 *For product characteristic x_{jk}, assume $\frac{\partial mc_j}{\partial x_{jk}} \geq 0$, $\frac{\partial q_j}{\partial x_{jk}} \geq 0$, and $\frac{\partial^2 q_j}{\partial x_{jk} \partial p_j} \geq 0$, with one inequality strict. Assume $q_j > 0$ and $\frac{\partial q_j}{\partial p_j} < 0$. Then a single-product firm will increase price if there is an increase in x_{jk}, holding all else constant.*

Proof

Differentiating the pricing rule with respect to x_{jk} leads to the comparative static

$$\frac{\partial p_j}{\partial x_{jk}} = \frac{\partial mc_j}{\partial x_{jk}} - \frac{\frac{\partial q_j}{\partial x_{jk}} \frac{\partial q_j}{\partial p_j} - \frac{\partial^2 q}{\partial p \partial x_{jk}} * q_j}{(\frac{\partial q_j}{\partial p_j})^2}. \tag{6}$$

The three terms in (6) are all non-negative, with at least one strictly positive. ||

By the first assumption, marginal costs of production are increasing in the amount of the characteristic. By the second assumption, demand for the good is increasing in the amount of the characteristic (consumer's value the characteristic). The third assumption requires demand to become less price elastic as the characteristic increases. When firms sell more than one product, the comparative statics are more complicated, as the expression in (6) contains additional terms that account for the effects on demand of the other products a firm sells. However, the comparative static will not equal zero unless all of these effects taken together completely offset one another.

Assume there exists a ξ_j that does affect utility for good j. Under mild regularity conditions, prices can be written as an implicit function of ξ and Z, where Z is defined as all other factors firms take into account at the time that firms make pricing decisions:

$$p_j = p(j, Z, \xi) \ \ j = 1, \ldots, J. \tag{7}$$

When ξ_j is omitted from the specification, the error term in utility is given by

$$e_{ij} = \beta_{i\xi}\xi_j + \epsilon_{ij}.$$

If sellers charge higher prices when their products have more desirable omitted characteristics, e_{ij} and p_j will be positively correlated after conditioning on the observed factors in Z.

When ϵ is assumed to be independent of price, so $f_\epsilon(\cdot)$ is the maintained distribution for the error, the direct effect of price on demand is negative, but the indirect effect via price's correlation with the omitted attribute is positive. The single price coefficient must account for both effects. When there is an omitted attribute the indirect effect leads to a finding that consumers are less sensitive to price than they actually are; they get more utility for paying the observed price than the econometrician has taken into account.

Addressing this problem directly in a likelihood setting is challenging when the omitted attribute is not observed. Letting e_i denote the vector $\langle e_{i1}, \ldots, e_{iJ} \rangle$ and $h_e(\cdot|p, Z)$ the density of e_i conditional on the observed prices. The choice probability for good j is obtained by integrating over $h_e(\cdot|p, Z)$ in (4) (instead of $f_\epsilon(\cdot)$) so one must be able to derive $h_e(\cdot|p, Z)$ from knowledge of (7) and other features of the model.

Two approaches that are similar in spirit to each other have been suggested as alternatives to the direct likelihood solution. One approach includes fixed-effects in the estimated specification for every product in every market to control for the unobserved factors, as in Chintagunta, Dube, and Goh (2005). The second approach, proposed in Berry (1994) and BLP matches observed to predicted shares to concentrate out product-market dummy variables.

Both of these approaches are also not without potential difficulties. They can substantially increase the dimension of the parameter space, making it computationally difficult to maximize a likelihood function that is rarely globally concave. Consistency of the estimator can be an issue; if the fixed effects are not consistently estimated because observations per product are not increasing as the sample size increases, the inconsistency is transmitted to the other estimated coefficients. A similar concern arises when shares are matched to concentrate out product-market dummy variables; because sampling error enters the objective function in a non-linear manner, the number of purchasers must grow at a rate substantially faster than the number of observed products.[6]

Given the difficulties associated with all three of these alternatives, in Sections 4-8 we develop tests that are powerful and easy to implement when looking for evidence of price endogeneity. A finding of "no evidence of price endogeneity" provides an argument against undertaking any of the three more difficult estimation approaches outlined above.

4 The Optimal Test for an Omitted Attribute

In these next two sections, we assume that the omitted variable is measured; the tests for this situation have been developed, and we review them to lay the foundation for their extension in section 6 to the case when the omitted variable is not observed.

For expositional convenience we start by assuming that the taste for the omitted attribute is common across individuals, so $\beta_{i\xi} = \beta_\xi$ (we loosen this assumption momentarily). From McFadden (1987) and Newey and McFadden (1994) we know the asymptotically optimal test for an omitted attribute in discrete choice differentiated product models is the likelihood ratio (LR) test of the null hypothesis $\beta_\xi = 0$. The test rejects with probability approaching one as the sample size increases. Additionally, against the (local) alternative $\beta_\xi = \frac{1}{\sqrt{N}}$, the test is asymptotically most powerful among all invariant tests.

Both the Lagrange Multiplier (LM) test and Wald test are asymptotically equivalent to the LR test. From a practical standpoint the LM test is useful because it does not require us to focus on estimation in the unconstrained

[6]See Berry Linton and Pakes (2004) who show that the rate conditions necessary for consistent estimation when observed shares are matched to predicted shares require J^2 observed purchasers when J products are in the choice set.

case; the LM test imposes $\beta_\xi = 0$ during estimation and then tests to see if the derivative of the likelihood with respect to β_ξ - evaluated at the constrained estimates - is large in absolute value. The details and intuition of the overidentification tests we develop are most clearly seen within the LM test setting. Later we turn to the Wald-type specification test.

Let $\tau = (\theta', \beta_\xi)'$ be the parameter vector with $R+1$ elements, with β_ξ the $R+1th$ element. The maximum likelihood estimator in the constrained case uses the sample analogs to the R population moments

$$E[\frac{\partial log P_i(\theta)}{\partial \theta'}' u_i(\theta)|X, p, \xi] = E[s_i(\theta)|X, p, \xi] \tag{8}$$

where $\frac{\partial log P_i(\theta)}{\partial \theta'}$ is the $J \times R$ matrix of derivatives evaluated at $\beta_\xi = 0$, $u_i(\theta) = Y_i - P_i(\theta)$, and $P_i(\theta) = (P_{i1}(\theta), \ldots, P_{iJ}(\theta))'$, and $s_i(\theta)$ denotes the $R \times 1$ score vector. The constrained parameter estimates $\hat{\theta}$ solve the sample analog equations

$$\frac{\partial log L_N(\hat{\theta})}{\partial \theta_r} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j} Y_{ij} \frac{1}{P_{ij}(\hat{\theta})} \frac{\partial P_{ij}(\hat{\theta})}{\partial \theta_r} = 0 \quad r = 1, \ldots, R. \tag{9}$$

When $\tau_0 = (\theta_0', 0)'$, $E[s_i(\theta_0)|X, p, \xi] = 0$, and $\hat{\theta}$ is consistent and asymptotically normal under standard regularity conditions (such as those given in Newey (1985)).

The LM test is based on the average value of the derivative of the likelihood with respect to β_ξ evaluated at the constrained parameter estimates:

$$\frac{\partial log L_N(\hat{\theta})}{\partial \beta_\xi} \;=\; \frac{1}{N} \sum_{i=1}^{N} \frac{\partial log P_i(\hat{\theta})}{\partial \beta_\xi}' u_i(\hat{\theta}). \tag{10}$$

Under H_0, $\frac{\partial log P_i(\theta_0)}{\partial \beta_\xi}$ is uncorrelated with $u_i(\theta_0)$, and $u_i(\theta_0)$ is mean zero, so the population analog to (10) is exactly equal to zero. The optimal statistic looks at the covariance between $\frac{\partial log P_i(\hat{\theta})}{\partial \beta_\xi}$ and $u_i(\hat{\theta})$, where

$$\frac{\partial log P_i(\hat{\theta})}{\partial \beta_\xi} = (\frac{\partial log P_{i1}(\hat{\theta})}{\partial U_{i1}} \xi_1, \ldots, \frac{\partial log P_{iJ}(\hat{\theta})}{\partial U_{iJ}} \xi_J)';$$

the omitted attribute scaled by a positive "hazard rate" of purchase with respect to utility. The test rejects if the omitted attribute can significantly account for any of the purchase decision residual.

The model can be extended to allow for heterogeneity in taste for the unobserved characteristic. In this general case the $R + L \times 1$ parameter vector is given by $\tau = (\theta', \beta_\xi')'$, with $\beta_\xi = (\beta_{\xi 1}, \ldots, \beta_{\xi L})'$ a $L \times 1$ parameter vector and $\beta_{i\xi} = \beta_{\xi 1} + \sum_{l=2}^{L} \beta_{\xi l} d_{il}$. This setting provides L moment conditions that under H_0 must tend to zero as the sample size increases. The sample analogs are given by

$$\frac{1}{N} \sum_i d_{il} * \frac{\partial log P_i(\hat{\theta})}{\partial \beta_{\xi l}}' u_i(\hat{\theta}) \quad l = 1, \ldots, L, \tag{11}$$

with $d_{i1} = 1 \; \forall i$ in the first moment. In the (most general) case, when the d_{il} are indicator variables that partition consumers into mutually exclusive and exhaustive types, these additional $L-1$ moment conditions look within each of the $L-1$ consumer types for covariance between the purchase residuals and the omitted attribute.

5 Conditional Moment Tests

This LM test can be viewed as a specific case of a wider class of available overidentification tests known as Conditional Moment tests (see Newey (1985) and Tauchen (1985)). The model is written as

$$Y_i = P_i(\theta_0) + u_i,$$

with

$$E[u_i(\theta_0) | Z, p, \xi, d_i] = 0. \tag{12}$$

This formulation makes it clear that any function of exogenous factors can be used in the construction of instruments for testing. In the case where L

separate instruments are available, we denote this $J \times L$ matrix for individual i as $g_i(Z, p, \xi, d_i) = (g_{i1}(\cdot), \ldots, g_{iL}(\cdot))$.

The direction(s) in which the test has power is determined by the choice of instrument(s). As Chamberlain (1987) shows, in a conditional moment framework the optimal instrument for testing whether any parameter's true value is non-zero is given by the expected value of the derivative of the error with respect to the parameter. For β_ξ, this is given by $E[\frac{\partial u_i(\tau_0)}{\partial \beta_\xi}|Z, p, \xi, d_i] = E[\frac{\partial P_i(\tau_0)}{\partial \beta_\xi}|Z, p, \xi, d_i]$. The LM test is the special case where the overidentification conditions use the instrument matrix: $g_i(Z, p, \xi, d_i) = \frac{\partial log P_i(\hat{\theta})}{\partial \beta_\xi}$.

A simpler-to-compute set of moments is available if the positive weights $\frac{\partial log P_{ij}(\hat{\theta})}{\partial U_{ij}} \forall j$ are dropped from (10) and (11). In this case for each i the $J \times L$ instrument matrix is

$$g_i = (\xi, d_{i2} * \xi, \ldots, d_{iL} * \xi),$$

and the sample analogs to the L population moments are given by

$$\overline{m}_N(\theta) = \frac{1}{N} \sum_i m_i(\theta) = \begin{bmatrix} \frac{1}{N} \sum_i \xi' u_i(\theta) \\ \frac{1}{N} \sum_i d_{i2} * \xi' u_i(\theta) \\ \cdots \\ \frac{1}{N} \sum_i d_{iL} * \xi' u_i(\theta) \end{bmatrix}. \tag{13}$$

They retain the covariance flavor of the optimal testing moments given in (10) and (11) and are equal to zero when evaluated at the true parameter vector. Under mild regularity conditions outlined in Newey (1985) (and reviewed in the Appendix), a central limit theorem can be applied to the sample mean $\sqrt{N}\overline{m}_N(\hat{\theta})$. When $plim\, \hat{\theta} = \theta_0$, its asymptotic variance is given by

$$\begin{aligned} \Omega(\theta_0) &= E[m_i(\theta_0) m_i(\theta_0)'] \\ &- E[m_i(\theta_0) s_i(\theta_0)'] * E[s_i(\theta_0) s_i(\theta_0)']^{-1} * E[m_i(\theta_0) s_i(\theta_0)']'. \end{aligned} \tag{14}$$

Lemma 2 *Assume conditions 1-8 from Newey (1985) hold. Let Ω_N denote any $L \times L$ positive definite symmetric matrix with $plim\, \Omega_N = \Omega$. Under H_0*

$$N * T_N(\hat{\theta}) = N * \ \overline{m}_N(\hat{\theta})' \, \Omega_N^{-1} \, \overline{m}_N(\hat{\theta}) \tag{15}$$

converges in distribution to a $\chi^2(L)$ *random variable.*

The proof is standard and uses the regularity conditions to establish that a Lindberg-Feller central limit theorem holds for $\sqrt{N}\Omega_N^{-1/2}\overline{m}_N(\hat{\theta})$.

The regularity conditions ensure that sample analogs to each component of (14) converge in probability to their population counterpart, so an estimator for Ω is readily available. However, Newey (1985) shows that $N * T_N(\hat{\theta})$ is easier to compute using standard regression packages. The uncentered r-squared from the regression of a vector of N ones on the $N \times R + L$ matrix given by stacking the N vectors $(m_i(\hat{\theta})', s_i(\hat{\theta})')$ is numerically equivalent to $N * T_N(\hat{\theta})$.[7]

While the power of any test must be determined on a case by case basis, for the logit distribution the conditional covariance test from (15) is the *optimal* test.

Lemma 3 *Assume* ϵ_{ij} *are i.i.d. type 1 extreme value. Then the covariance test provided by (15) is numerically equivalent to the LM test, and is therefore consistent and asymptotically (locally) most powerful.*

Proof

$\hat{\theta}$ solves (9). In the type 1 extreme value case

$$\begin{aligned} \frac{\partial log L_N(\hat{\theta})}{\partial \theta_r} &= \frac{1}{N}\sum_{i=1}^{N}(\sum_j Y_{ij}x_{jr} - (\sum_j P_{ij}(\hat{\theta})x_{jr})) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_j (Y_{ij} - P_{ij}(\hat{\theta}))x_{jr} = 0. \end{aligned} \tag{17}$$

[7]Identification ensures that these conditional covariance tests are consistent except under a very special circumstance. Let $P_j^* = \frac{1}{N}\sum_i Y_{ij}$ and $\overline{\theta} = plim\,\hat{\theta}$ under the alternative. Identification requires $P_j^* \neq P_j(\overline{\theta})$ for at least two choices. This leads to an asymptotic value for the moment conditions from (13) to be different from zero unless these differences in product specific probabilities average out across products. Alternatively, the test is consistent unless $\forall l$

$$\sum_j P_{lj}(\overline{\theta})\xi_j - \sum_j P_{lj}^*\xi_j = 0 \tag{16}$$

||

Additionally, the covariance approach to testing in linear settings is very often optimal, although one must be cautious about drawing analogies between linear and non-linear settings. For these reasons we believe this conditional covariance test should have reasonable power in many settings.

6 Testing when the Attribute is Unobserved

In most cases an attribute is omitted because it is not observed. In this section we develop a proxy for the unobserved attribute based on its covariance with price conditional on Z, denoted $COV(p, \xi|Z)$ (see Section 2). We use this proxy as the instrument for testing in the conditional moment formulation given by (13).

When prices are endogenous p is not independent of ξ conditional on Z. Assuming the dependence is reflected in a non-zero conditional covariance between p and ξ, so $COV(p, \xi|Z) \neq 0$, there exists a (γ_0, γ_1) and $\nu_j, j = 1, \ldots, J$ such that

$$p_j - E[p_j|Z] = \gamma_0 + \gamma_1(\xi_j - E[\xi_j|Z]) + \nu_j \tag{18}$$

with $\gamma_1 \neq 0$, $\sum_j \nu_j = 0$, and $\sum_j \xi_j \nu_j = 0$. $p_j - E[p_j|Z]$ is a valid instrument for the test as it covaries with ξ_j. However, it is not observed because $E[p_j|Z]$ is not known. We estimate $E[p_j|Z]$, described in section 8, and define the proxy $\hat{\xi}_j$ as the difference between p_j and the estimate $\hat{E}[p_j|Z]$:

$$\hat{\xi}_j = p_j - \hat{E}[p_j|Z] = \gamma_0 + \gamma_1(\xi_j - E[\xi_j|Z]) + \nu_j + \eta_j, \tag{19}$$

where $\eta_j = E[p_j|Z] - \hat{E}[p_j|Z]$ denotes the estimation error.

Under standard regularity conditions $\hat{E}[p_j|Z]$ will converge to $E[p_j|Z]$ as the sample size increases. We note that our results do not require that this condition hold. Under the null hypothesis, the η_j add an additional source of variance to the moments, but this is correctly accounted for in the size of the test. Under the alternative hypothesis, the test remains consistent as long as the conditions outlined in Lemma 6 hold.

We replace ξ_j with $\hat{\xi}_j$ in the sample moments:

$$\hat{\bar{m}}_N(\theta) = \frac{1}{N}\sum_i \hat{m}_i(\theta) = \begin{bmatrix} \frac{1}{N}\sum_i \hat{\xi}' u_i(\theta) \\ \frac{1}{N}\sum_i d_{i2} * \hat{\xi}' u_i(\theta) \\ \dots \\ \frac{1}{N}\sum_i d_{iL} * \hat{\xi}' u_i(\theta) \end{bmatrix}, \tag{20}$$

where $\hat{m}_i(\theta)$ denotes the use of $\hat{\xi}$. When $\tau_0 = (\theta_0, 0)$, $plim\, \hat{\theta} = \theta_0$ and

$$\begin{aligned} \hat{\Omega}(\theta_0) &= E[\hat{m}_i(\theta_0)\hat{m}_i(\theta_0)'] \\ &- E[\hat{m}_i(\theta_0)s_i(\theta_0)'] * E[s_i(\theta_0)s_i(\theta_0)']^{-1} * E[\hat{m}_i(\theta_0)s_i(\theta_0)']'. \end{aligned} \tag{21}$$

The new test statistic is:

$$N * \hat{T}_N(\hat{\theta}) = N * \ \hat{\bar{m}}_N(\hat{\theta})' \hat{\Omega}_N^{-1} \hat{\bar{m}}_N(\hat{\theta}),$$

where $\hat{\Omega}_N$ is constructed such that $plim\ \hat{\Omega}_N = \hat{\Omega}(\theta_0)$.

These moment conditions differ from (13). For $N * \hat{T}_N(\hat{\theta})$ to be a valid test statistic one must (re)establish in this setting that Assumptions 1 and 3-8 from Newey (1985) still hold under the null hypothesis. Assumptions 1 and 3-6 are model specification, regularity, and identification conditions, and are typically immediately verifiable (see Appendix). We investigate Assumptions 7 and 8 further.

Assumption 7 states that the conditional moments used for testing have zero mean when the true parameter value $\tau_0 = (\theta_0, 0)$ and the moment is evaluated at θ_0. Lemma 4 establishes this result.

Lemma 4 *If* $\tau_0 = (\theta_0, 0)$ *then* $E[\hat{\bar{m}}(\theta_0)|Z, p] = 0$.

Proof

Consider an arbitrary moment l from $\hat{m}_i(\theta_0)$:

$$
\begin{aligned}
E[\hat{m}_{il}(\theta_0)|Z,p,d_i] &= E[d_{il} * \hat{\xi}' u_i(\theta_0)|Z,p,d_i] \\
&= d_{il} * \hat{\xi}' E[u_i(\theta_0)|Z,p,d_i] \\
&= d_{il} * \hat{\xi}'(E[Y_i|Z,p,d_i,\theta_0] - P_i(\theta_0)) \\
&= 0
\end{aligned} \tag{22}
$$

where the second equality follows from the fact that $\tau_0 = (\theta_0, 0)$ and $\hat{\xi}$ is in the conditioning set because it is a function of p and Z, and the final equality holds because $E[Y_i \mid Z, p, d_i, \theta_0] = P_i(\theta_0)$ (from 12). Both i and l are arbitrary, so the expectation is zero $\forall i$ and $\forall l$, and thus also in the aggregate. $\|$

Assumption 8 requires the testing moments $\hat{m}_i(\theta)$ to be linearly independent of the score $s_i(\theta)$. The practical requirement is that there exists an observable factor that is correlated with ξ_j but not perfectly collinear with (X_j, p_j), the arguments that enter into utility for good j. Our discussion in Section 2 pointed to $p_j - E[p_j|Z]$ as a possible proxy, where other exogenous cost and demand factors that cause price to vary but do not enter utility for good j are included in the conditioning set Z when estimating $\hat{E}[p_j|Z]$.

Lemma 5 *Assume conditions 1 and 3-8 from Newey (1985) hold. Let $\hat{\Omega}_N$ denote any $L \times L$ positive definite symmetric matrix with plim $\hat{\Omega}_N = \hat{\Omega}$. If $\tau_0 = (\theta_0, 0)$ then $N * \hat{T}_N(\hat{\theta})$ converges in distribution to a $\chi^2(L)$ random variable.*

This is a second application of a Lindberg-Feller central limit theorem to $\sqrt{N}\hat{\Omega}_N^{-1/2}\hat{\overline{m}}_N(\hat{\theta})$.

It remains for us to show the conditions under which the test is consistent. Under the alternative, $plim\, \hat{\tau} = \overline{\tau} = (\overline{\theta}, 0) \neq \tau_0$. The test is consistent if $plim\, \hat{T}_N(\hat{\theta}) = \hat{T}(\overline{\theta}) > 0$, which occurs if and only if one or more of the L

testing moments have non-zero expectation. Lemma 6 provides these conditions; notationally we let $\nu = (\nu_1, \ldots, \nu_J)$ and $\eta = (\eta_1, \ldots, \eta_J)$, and we treat the d_l's as group indicator variables, where $d_{i1} = 1 \, \forall i$.[8]

Lemma 6 *Let the covariance between ξ and $u_i(\overline{\theta})$ for subgroup "d_l" be given by*

$$K_l = E[d_{il} * (\xi - E[\xi|Z])' u_i(\overline{\theta})|Z, p]$$

and let

$$A_l = E[d_{il} * (\nu + \eta)' u_i(\overline{\theta})|Z, p].$$

Assume $|A_l| < \infty$ and $|K_l| < \infty \; \forall l$. Assume $COV(p, \xi|Z) \neq 0$ and $\hat{\Omega}(\overline{\theta})$ is positive definite. If $\gamma_1 K_l + A_l \neq 0$ for at least one l, then plim $\hat{T}_N(\hat{\theta}) = \hat{T}(\overline{\theta}) > 0$.

Proof

$$\begin{aligned} E[d_{il} * \hat{\xi}' u_i(\overline{\theta})|Z, p] &= E[d_{il} * \gamma_0 * \textstyle\sum_j u_{ij}(\overline{\theta})|Z, p] \\ &+ E[d_{il} * \gamma_1 * (\xi - E[\xi|Z])' u_i(\overline{\theta})|Z, p] \\ &+ E[d_{il} * (\nu + \eta)' u_i(\overline{\theta})|Z, p]. \end{aligned} \tag{23}$$

$u_{ij}(\theta)$ is the residual, so by construction $\sum_j u_{ij}(\theta) = 0 \, \forall i$ and $\forall \theta$. The second term equals $\gamma_1 K_l$. The third term equals A_l. The expected value of the moment is non-zero if $\gamma_1 K_l + A_l \neq 0$. $\|$

$\gamma_1 \neq 0$ because $COV(p, \xi|Z) \neq 0$. The test is consistent if one or more of these L inequalities holds, that is, if $\gamma_1 K_l \neq -A_l$ for at least one l.

7 The Wald Formulation

An alternative to the conditional moment test is the Wald test, which is also a generalized method of moments overidentification test. This test includes the

[8] The Lemma is easily modified if d_l is not an indicator variable.

proxy directly in the estimated specification. Letting λ denote the parameter on the proxy, utility is specified as

$$U_{ij} = \sum_k x_{jk}\beta_{ik} - \alpha_i p_j + \lambda\hat{\xi}_j + \epsilon_{ij}, \tag{24}$$

with $\lambda\hat{\xi}_j$ as the control function. Under the null hypothesis of exogenous prices the coefficient on the proxy is equal to zero. The specification can be extended to include a full set of L interactions between demographics d_i and the proxy; if any of the additional L coefficients enters significantly, the null hypothesis is rejected.

Under the null hypothesis both the conditional moment and the Wald tests utilize moments that have population analogs equal to zero when evaluated at θ_0. The entire set of moments used for estimation and testing is similar across both approaches, and the tests would be asymptotically equivalent but for the fact that $\hat{\xi}_j$ is used instead of ξ_j. When d_i is not interacted with $\hat{\xi}_j$, the sample $R+1$ moments that are zeroed under the Wald approach are given as

$$\hat{E}[\frac{\partial log P_i(\hat{\theta},\hat{\lambda})}{\partial(\theta',\lambda)}' u_i(\hat{\theta},\hat{\lambda})|X,p,\hat{\xi}] = 0. \tag{25}$$

The test is consistent as long as $plim\,(\hat{\theta},\hat{\lambda}) \neq (\overline{\theta},0)$ when the null is false.

The Wald approach is attractive for three reasons. The test statistic is even easier to compute than the conditional moment test, requiring no additional calculation beyond standard likelihood function output; significance of the coefficient is equivalent to rejecting price exogeneity (when $\tau_0 = (\theta_0, 0)$ the size of the test is not affected by the first stage estimation of $E[p_j|Z]$). The resulting estimates are suggestive of the direction and magnitude of the bias arising from the price endogeneity, although one must be cautious given that $\hat{\lambda}\hat{\xi}_j \neq \xi_j$ generally. Finally, the point estimates and standard errors can be used to construct ranges for starting values for any one of the three more complicated correction approaches suggested in Section 3. Good starting values can be helpful when the objective function is non-linear in parameters, there are a large number of parameters, and/or there are possibly multiple local maxima, which are often characteristics of these three approaches.

Two related issues arise with specification and inference with respect to these last two objectives. First, unless $\hat{\lambda}\hat{\xi}_j = \xi_j \forall j$, some discrepancy from measurement error will arise which leads to the true residual $u_{ij}(\xi)$ to differ from the model prediction $u_{ij}(\hat{\xi})$. Given the non-linear environment, one may want to include a random effect to account in part for this error, especially if its variance is large. We estimate the specification given by

$$U_{ij} = \sum_k x_{jk}\beta_{ik} - \alpha_i p_j + \lambda\hat{\xi}_j + \sigma\iota_j + \epsilon_{ij}, \tag{26}$$

where σ is a new parameter and ι_j is distributed $N(0,1)$. Second, when $\tau_0 \neq (\theta_0, 0)$ and one wants to construct a range of possible parameter values under this alternative, the source of error from the first stage estimation of $E[p_j|Z]$ should be taken into account. We bootstrap to account for this additional source of error, as we describe in our applications later.

8 Estimation of the proxy/instrument

The purpose of estimating $E[p_j|Z]$ is to recover the proxy $\hat{\xi}_j = p_j - \hat{E}[p_j|Z]$. As mentioned earlier, $\hat{\xi}_j$ cannot be perfectly collinear with the score moments or the test has no power, so a variable excluded from X_j that enters the pricing function for j must be observed. The power of the test is increasing in $COV(\xi, \hat{\xi})$, which is determined by the variation that is used to estimate $E[p_j|Z]$ and the restrictions placed on $E[p_j|Z]$ during estimation. This variation is likely to come from one or more of the following sources: within market at a given time, across markets at a given time, or within/across markets over time. To make the discussion concrete, we illustrate specification and estimation of $E[p_j|Z]$ with recent demand applications that make use of different sources of variation: Goolsbee and Petrin (2004), who look at cable and satellite television demand, Berry, Levinsohn, and Pakes (1995), who look at automobiles, and Chintagunta, Dube, and Goh (2005), who look at margarine.

In Goolsbee and Petrin (2004), almost 30,000 households are observed in over 300 geographically distinct television markets, and four alternatives are available to households in every market: (1) antenna only, (2) expanded basic cable service, (3) expanded basic cable with a premium service added, such as HBO, and (4) satellite dish. The price endogeneity problem arises because unobserved factors like service or average channel quality are correlated with price but not observed by the authors. With over 300 variants of each type of product, e.g. expanded basic cable, one can separately estimate $E[p|j, Z]$ for each product using the cross-market variation (allowing the coefficients of each function to differ for each type of product). Z includes X_j and may include all of the other characteristics of other products in the market, and any other relevant demand and cost factors.

Berry, Levinsohn, and Pakes (1995) observe over 100 market-level observations on prices, quantities, and characteristics of automobiles sold in the U.S. for every year from 1971 to 1990. The price endogeneity problem arises because, with only five included characteristics, additional unobserved quality (e.g.) is correlated with price. With the automobile data, very few observations are available on the same nameplate (i.e. the same product) over time, because cars enter, exit, and change quickly. Unlike the television case, this means some restrictions on $E[\, p_j \mid Z]$ across vehicles will be necessary. Some possibilities include restricting parameters to be the same across: all products and all years, all products within the same year, similar products within a year, or similar products across years.

A second difference with the television case is that the number of potential arguments entering $E[\, p_j \mid Z]$ may be quite large; in one extreme case, every product's characteristics may affect every product's price. In a multi-product multi-competitor market Pakes (1994) suggests a parsimonious approach that includes three regressors for each characteristic: the characteristic itself, the sum of the characteristic across own-firm products (excluding that product), and the sum of the characteristic across rival firm products.

Chintagunta, Dube, and Goh (2005) (CDG) observe weekly purchase histories of 992 households between January 1993 and March 1995 collected using checkout-counter scanners. The market for margarine is similar to that for television in the sense that there are only four choices in their data: Blue Bonnet, I Can't Believe It's Not Butter (ICBINB), Parkay, and Shedd's. However, unlike television demand, the physical characteristics of the products are not changing over time. Instead, the price endogeneity problem arises because weekly retail prices covary with demand-shifting marketing-mix variables which may not be observed by the econometrician, including: whether the product is on display, whether it is featured, changes in its shelf-space, the availability of coupons (in-store or not), or promotions in complementary or substitute categories. CDG observe wholesale prices, which affect retail prices but do not enter into consumer utility conditional on the retail price. Thus one could use the residual from the regression of product's retail price at time t on an intercept and its wholesale price at time t. If $COV(p, \xi|Z) > 0$, a large residual is suggestive of more marketing-mix activities.

9 Demand for Television

Our empirical application applies the test to the model of demand for households' choice among television reception options from Goolsbee and Petrin (2004), who emphasize the importance of omitted attributes. We estimate both the constrained model (without the attribute) and the Wald-type test specification described in Section 7, where $\hat{\xi}$ are included directly as new regressors in the likelihood function.

9.1 Data and Demand Specification

The data and specification are very similar to Goolsbee and Petrin (2004). The data come from two sources: Forrester Technographics 2001 and Warren Publishing's 2001 Cable and Television Factbook. We restrict our analysis to a subsample of the 30,000 households from the original work, using 11,810 households that reside in 172 geographically distinct television markets. Each market contains only one cable franchise, and four alternatives are available to households: (1) antenna only, (2) expanded basic cable service, (3) expanded basic cable with a premium service added, such as HBO, and (4) satellite dish.

Estimated utility is given as

$$U_{ij} = \alpha p_{mj} + \sum_{g=2}^{5} \theta_g p_{mj} 1_{ig} + \beta_0' x_{mj} + \gamma_j' d_i + \sigma \iota_i c_j + \lambda_j \hat{\xi}_{mj} + \epsilon_{ij}. \tag{27}$$

x_{mj} are the observed characteristics of the product and includes a product intercept term. $\hat{\xi}_{mj}$ is the proxy and has λ_j as its coefficient. 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.[9] 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: $c_j = 1$ if j is one of the three non-antenna alternatives and $c_j = 0$ otherwise, ι_i is an i.i.d. standard normal deviate,

[9] 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 σ is its standard deviation, reflecting the degree of correlation among the non-antenna alternatives. ϵ_{ij} is i.i.d. extreme value.[10]

As mentioned earlier, when one is interested in learning about the magnitude of the bias and/or constructing intervals of reasonable starting values, it is useful to include a random effect to account in part for the difference between ξ and $\hat{\xi}$. Here we do not try to separately identify the variance of this random effect from the variance of normal deviate used for correlation in taste across the three multi-channel video options.

In the Forrester survey, respondents reported the type of television medium that they have.[11] The Forrester survey also provides the demographic information they use, including family income, household size, education, and type of living accommodations. Finally, the survey includes an identifier for the household's television market, which links households 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 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, one for each of the two cable choices for

[10]The error specification in Goolsbee and Petrin (2004) is more flexible; they use a multivariate normal specification in place of the logit error.

[11]Specifically, they report whether they have cable or satellite, and the amount they spend on premium television. Respondents are classified 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.

each operator. Satellite prices do not vary geographically, and the price of antenna-only is assumed to be zero.[12] More complete details are provided in Goolsbee and Petrin (2004).

9.2 Estimation of $\hat{\xi}_{mj} = p_{mj} - E[p_{mj}|Z]$

We estimate the expected price functions product by product using the cross-market variation from the 172 different observations on each product. However, since price does not vary across geographic location for antenna-only and satellite, we do not construct proxies for these products. We obtain the proxy for expanded basic by regressing its price on all of the product attributes listed above for the product choices available in the market. In addition we include Hausman (1997)-type price instruments, one for expanded basic and premium each. 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, and is intended to reflect common costs of the multiple system operator. The premium proxy is constructed in a similar manner.

9.3 Results

Table 1 gives the estimated parameters and standard errors for the two approaches. While not necessary for testing, we bootstrap to account for the additional variance from estimating the expected price (see Section 7) because we are interested in constructing approximations for the range for the bias suggested by the Wald-type test specification.[13]

[12]For the price of satellite, we use \$50 per month plus an annual \$100 installation and equipment cost.

[13]To approximate the additional source of variance arising from the estimated expectation, we add a new term to the standard estimate of variance of the parameters. We calculate the new term by first drawing a bootstrapped sample of prices and observed de-

Table 1

TV Reception Choice

Constrained Approach and the Wald-type Test Specification

Alternatives: 1. Antenna only, 2. Expanded Basic cable, 3. Premium cable, 4. Satellite Dish

Variables enter alternatives in parentheses and are zero in other alternatives.

Explanatory variable	Constrained Approach	Wald-type Test Specification
	(Standard errors in parentheses)	
Price, in dollars per month (1-4)	-.0202 (.0047)	-.0969 (.0400)
Proxy for expanded-basic cable price (2)		.0805 (.0416)
Proxy for premium cable price (3)		.0873 (.0418)
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)
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)
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		

For each of the seven largest multiple system operators we include separate indicators for expanded basic and for premium.

Table 2

Television Choice Elasticities:

Constrained Approach and Wald-type Test Specification

	Constrained Approach	Wald-type Test Specification
Price of expanded-basic cable		
Antenna-only share	Upward	0.96
Expanded-basic cable share	Sloping	-1.18
Premium cable share	Demands	0.99
Satellite share	—	0.95
Price of premium cable	—	
Antenna-only share	—	0.60
Expanded-basic cable share	—	0.65
Premium cable share	—	-2.36
Satellite share	—	0.64
Price of satellite	—	
Antenna-only share	—	0.43
Expanded-basic cable share	—	0.48
Premium cable share	—	0.48
Satellite share	—	-3.79

The first column of Table 1 gives the constrained model while the second column includes $\hat{\xi}_{mj}$ as new regressors. The base price coefficient is reported first, followed by the coefficients on the expanded basic and premium proxies. In the constrained model 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. It is also highly significant. Inclusion of the proxies raises the magnitude of the estimated base price coefficient by 500% and the standard error by 1000%. Both proxies enter significantly and with a positive sign, identifying the price endogeneity problem and suggesting that products with large proxies possess desirable attributes omitted from the specification.

Table 2 reports the implied price elasticities computed from the constrained model parameter estimates and from those estimates with the bias correction suggested by the Wald-type test specification. The correlation between price and the unobserved characteristic is so strong that demands are upward sloping in the constrained model. When the predicted bias in the parameter estimates from the Wald-type test is "added back" to the constrained elasticity estimate, one obtains the elasticities in column 2. They are much more reasonable, as all own-price elasticities are greater than 1, which is in accordance with static pricing theory.

In Petrin and Train (2004) we explore the ability of the control function approach to estimate the magnitude of the bias arising from endogenous

mand and supply factors from the 172 markets. We then estimate $\hat{E}[p|j, Z]$ and calculate the implied (new) proxies, and then re-estimate the model with these proxies. We repeat this process and then compute the the variance in parameter estimates over the bootstrapped price samples, adding this variance to the traditional formulas. 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%. 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.

prices in the automobile demand paper of BLP and the margarine demand paper of CDG. There we report that use of the simplest proxy formulations described in Section 8 - when added to their constrained specifications - reject exogenous prices at any standard level of statistical significance. The control function setup also provides a close estimate of the magnitude of the bias. Interested readers are referred to this paper for more details.

10 Monte Carlo Experiments

We construct Monte Carlo data for different situations with an unobserved attribute correlated with price. A product is sold in each of several markets and its attributes and price vary by markets. Each consumer lives in one market and either buys or does not buy the product offered in that market. The utility that consumer i who lives in market m obtains from the product is

$$U_i = \beta_0 + \beta_i x_m + \beta_\xi \xi_m - \alpha p_m + \epsilon_i,$$

where x_m is a product attribute that is observed by the researcher, ξ_m is a product attribute that is not observed by the researcher, and p_m is the price of the product. β_i, the random coefficient on x_m, is distributed $N(\beta_x, \sigma_x^2)$, and ϵ_i is distributed logit. Given these assumptions the share of consumers buying the product in market m is given by

$$s_m = \int \frac{exp(\beta_0 + \beta_i x_m + \beta_\xi \xi_m - \alpha p_m)}{1 + exp(\beta_0 + \beta_i x_m + \beta_\xi \xi_m - \alpha p_m)} dP(\beta_i). \tag{28}$$

We examine four different monte carlo cases that are designed to generate a wide range of correlations between price and the unobserved factor while maintaining a reasonable range of prices and market shares. Price is set at a markup over marginal cost based on static profit maximization:

$$p_m = mc_m - \frac{s_m}{\partial s_m / \partial p_m}.$$

We vary the specification for marginal cost across the cases, which gives rise to the variation in $Corr(p_m, \xi_m)$. In all specifications x_m enters marginal cost, as do two other variables which do not affect demand: w_m, which is observed by the researcher, and a_m which is not. In two of the four specifications ξ_m does not enter marginal costs, which unambiguously lowers $Corr(p_m, \xi_m)$. In two cases cost factors enter marginal costs linearly, and in two cases they enter exponentially.

Each of the random variables x_m, ξ_m, w_m, and a_m is assumed to be an i.i.d. N(0,0.5) deviate. All parameters of the utility and cost functions are set equal to 1, except for the intercepts, which equal 10, and the variance in taste for x_m, which is equal to 0.5. Cases 1-4, in order, specify marginal cost as follows:

$$
\begin{aligned}
mc_m &= 10 + x_m + \xi_m + w_m + a_m \\
mc_m &= 10 + x_m + w_m + a_m \\
mc_m &= 10 + exp(x_m + \xi_m + w_m + a_m) \\
mc_m &= 10 + exp(x_m + w_m + a_m).
\end{aligned}
$$

The researcher is assumed to observe 20 purchase decisions per market for 200 markets, which are used to construct the estimate for s_m. The researcher does not observe ξ_m and a_m in each market, instead seeing only x_m, w_m, and p_m, and OLS is used to estimate the residuals $\hat{\xi}_m = p_m - \hat{E}[p|x_m, w_m]$. The demand parameters are estimated using maximum likelihood in two ways. The constrained case excludes $\hat{\xi}$ from the specification to see how much bias the price endogeneity generates. The Wald approach from Section 7 includes $\hat{\xi}$ directly in the likelihood function and tests its coefficient λ for significance.

We want to explore how well the Wald-type testing framework is able to identify the magnitude of the bias caused by the endogeneity. With this goal in mind, we undertake the two additional steps suggested in Section 7: we add a random effect to the specification to (in part) reflect the discrepancy between ξ_m and $\hat{\lambda}\hat{\xi}_m$, and we account for the sampling error arising because $E[p|x_m, w_m]$ is estimated. The former adds the new parameter σ, the variance of a random effect, and leads to estimation of the utility function given by

$$\tilde{U}_i = \beta_0 + \beta_i x_m + \lambda\hat{\xi}_m - \alpha p_m + \sigma\iota + \epsilon_i,$$

where ι is distributed $N(0,1)$. For the latter we bootstrap, repeating estimation 500 times (that is, on 500 different datasets, each with 200 markets) and use the variance of the parameter estimates across these 500 cases to construct estimates of the distribution of parameters under the null and alternative hypotheses.

Table 3 summarizes the Monte Carlo data. Reported statistics are calculated for each sample of 200 markets, and then are averaged over the 500 iterations. For all four specifications prices range from approximately 10 to 14. For the linear marginal cost specifications shares range from 0.11 to 0.34. For the exponential specifications they range from 0.01 to 0.16, with over 50% of markets having less than 10% of consumers purchasing. $Corr(p_m, \xi_m)$ is respectively 0.55, 0.15, 0.40, and 0.03, and is substantially higher in specifications 1 and 3 where the unobserved demand attribute affects costs.

Table 4 reports the parameter estimates for the constrained approach and for the Wald-type test specification from Section 7. The average of the parameter estimates and their standard deviation across the 500 iterations is reported first for the constrained logit model. Next, the six parameters associated with the Wald test specification are reported, where the two new parameters are σ, the standard deviation of the random effect, and λ, the coefficient on $\hat{\xi}_m$. Test results for the null hypothesis that $\lambda = 0$ are reported at the bottom. We discuss each case in turn.

Table 3

Summary of Data Generated in Monte Carlo Samples

True values: $\alpha = 1$, $\beta_0 = 10$, $\beta_x = 1$, $\sigma_x = 0.5$				
Case	1	2	3	4
Price Range				
Mean	11.31	11.33	12.75	12.57
10%-90%	10.17-12.45	10.36-12.28	11.41-14.60	11.49-14.03
Share Range				
Mean	0.22	0.23	0.09	0.10
10%-90%	0.11-0.34	0.10-0.38	0.01-0.16	0.02-0.19
% of markets with				
shares < 0.01	0.0%	0.0%	7.4%	5.7%
shares < 0.10	5.8%	10.0%	57.1%	53.5%
Corr(p_m, ξ_m)	0.55	0.15	0.40	0.03
Marginal Costs				
Linear	Yes	Yes	No	No
Expontial	No	No	Yes	Yes
Include ξ_m	Yes	No	Yes	No

Reported numbers are the statistics for each of the samples with 200 markets averaged over the 500 Monte Carlo iterations.

10.1 Case 1: Linear marginal costs

Case 1 has marginal costs linear in all of the demand and cost factors. The average correlation between price and the unobserved attribute is 0.55. The constrained approach is severely biased downward, with the price coefficient estimated to be 0.51, half of the true value of 1 (the standard deviation is 0.03). The other estimates are similarly biased down to almost half of their true values.

When $\hat{\xi}$ is included in the specification, λ enters with a coefficient of 0.62 and a standard deviation of 0.10. The test rejects the null hypothesis in every one of the 500 replications at a size of 0.01. The average estimate of the price coefficients across the iterations is equal to 0.99, with a standard deviation 0.09. The other averaged point estimates are similarly close to their true values. Thus, the Wald-type test identifies the specification problem in every case and almost perfectly predicts the magnitude of the bias.

10.2 Case 2: Unobserved demand attribute does not affect (linear) marginal cost.

Case 2 has linear marginal costs, but ξ_m excluded, lowering the correlation between price and ξ_m to 0.15. The constrained model continues to perform poorly, with the point estimate for the price coefficient on average equal to 0.74 (with standard deviation 0.06). Other parameter estimates are similarly biased down. λ enters with a coefficient that is on average equal to 0.39, and the test rejects in over 99% of the 500 cases at the 0.01 significance level. The magnitude of the bias is identified again, as the estimated price coefficient is equal to 1.01, with standard deviation 0.26, and the other coefficient estimates are also close to their true values: 10.13 vs. 10, 1.01 vs. 1, and 0.52 vs. 0.5.

Table 4

Constrained Model and Wald-type Test Specification for Monte Carlo Data

200 markets, 20 observations per market, 500 iterations

True values: $\alpha = 1$, $\beta_0 = 10$, $\beta_x = 1$, $\sigma_x = 0.5$				
Case	1	2	3	4
Constrained Model				
α	0.51	0.74	0.63	0.86
	(0.03)	(0.06)	(0.03)	(0.05)
β_0	4.56	7.17	5.52	8.40
	(0.38)	(0.69)	(0.43)	(0.66)
β_x	0.51	0.74	0.65	0.88
	(0.06)	(0.09)	(0.07)	(0.09)
σ_x	0.23	0.24	0.12	0.22
	(0.45)	(0.49)	(0.33)	(0.41)
Wald Test/Control Function				
α	0.99	1.01	0.81	0.94
	(0.09)	(0.26)	(0.08)	(0.09)
β_0	9.90	10.13	7.69	9.20
	(1.11)	(2.48)	(0.95)	(1.02)
β_x	0.99	1.01	0.99	1.00
	(0.12)	(0.25)	(0.13)	(0.14)
σ_x	0.44	0.52	0.26	0.41
	(0.23)	(0.40)	(0.19)	(0.25)
σ	0.28	0.53	0.55	0.60
	(0.20)	(0.69)	(0.35)	(0.45)
Test statistic: λ	0.62	0.39	0.25	0.08
	(0.10)	(0.15)	(0.07)	(0.07)
Rejection rate at:				
size=0.01	100%	99.2%	99.6%	85.2%
size=0.10	100%	99.4%	99.6%	93.2%

Average parameter estimate and standard deviation (in parentheses) across 500 iterations. For the size of tests, 500 iterations are used to estimate the distribution of λ under the null hypothesis (estimates not reported here).

10.3 Case 3: Exponential marginal costs

Case 3 defines marginal cost to be exponential in all demand and cost factors. Since the cost variables do not enter utility, in some markets the high prices lead to very low demand; on average 7% of markets have demand shares less than 1%. The non-linearity causes the correlation between price and the unobserved demand attribute to fall to 0.40 (relative to the linear case of 0.55).

The constrained estimates are again severely biased downward. The average point estimate on price is 0.63 and the standard deviation is 0.03, and other coefficients exhibit similar problems. The Wald-type test again easily identifies the specification problem, rejecting at 0.01 significance level in over 99% of cases. The Wald estimate for the price coefficient is on average 0.81 with standard deviation 0.09, so while the approach points correctly to the direction of the bias, it appears to only identify about half of the bias in the estimated price coefficient (0.81 vs. 0.63).

10.4 Case 4: Unobserved demand attribute does not affect (exponential) marginal cost.

Case 4 defines marginal cost to be exponential in demand and cost factors with ξ_m excluded. With non-linearity in costs this exclusion causes the correlation between price and the unobserved demand attribute falls to 0.03.

The price coefficient in the constrained approach is again biased down. The test continues to identify the unobserved attribute problem, rejecting at significance level 0.01 in 85% of cases and at 0.10 level 93% of cases. The Wald-test specification again identifies the magnitude of the bias, as the coefficients are on average close to the truth; the price coefficient is 0.94 vs. 1, and the other coefficients are respectively 9.2 vs 10, 1.0 vs 1.0, and 0.41 vs 0.5.

11 Conclusion

In applications of differentiated product models all of the relevant product attributes may not be observed by the econometrician. In this case price may be positively correlated with the unobserved portion of consumers' utility for that product: producers charge more and consumers are willing pay more for products with more of the omitted attribute, holding all else constant. This positive correlation biases estimates of price elasticities towards zero, and evidence of it has been found in many applications spanning a wide range of markets and differing data types.

In this paper we develop a class of overidentification tests that are easy to implement in standard regression packages and have reasonable power against the "exogenous prices" hypothesis. The tests include both conditional moment tests and control function type tests of specification. The tests reject if the demand residuals covary with the proxies for the unobserved attribute, conditional on the other factors in the demand model.

A major advantage of our testing approaches is their simplicity. They all use Ordinary Least Squares in an initial stage to obtain the proxies, followed by a second stage of likelihood function maximization. The control function test includes the proxies directly in the maximization of the likelihood function and tests for their significance. The conditional moment test estimates the likelihood model without the proxies, and then tests for correlation between the proxies and the (generalized) demand residuals.

We describe approaches to implementing the tests for three of the recent demand exercises mentioned above, where prices are shown to be endogenous. Each of these exercises uses a different kind of data, including aggregate (market-level) data (Berry, Levinsohn, and Pakes (1995)), household-level cross-sectional data (Goolsbee and Petrin (2004)), and household-level panel data (Chintagunta, Dube, and Goh (2005)). Our Monte Carlos draw on a formulation where prices are set according to a non-linear in characteristics, inverse-elasticity rule. Our simplest test specifications identify the price endogeneity problem in every case and provide a close estimate of the magnitude of the bias from omitting the attribute.

References

Bajari, P., and C. Benkard (2003): "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," Working Paper, Department of Economics, Stanford University.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–889.

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

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

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–344.

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.

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.

HECKMAN, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

HECKMAN, J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impacts of Interventions: An Overview," *Journal of Econometrics*, 30, 239–267.

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.

MCFADDEN, D. (1987): "Regression-Based Specification Tests for the Multinomial Logit Model," *Journal of Econometrics*, 34, 63–82.

MCFADDEN, D., AND W. NEWEY (1994): *Large Sample Estimation and Hypothesis Testing*Elsevier, chap. 36.

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

NEWEY, W. K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53(5), 1047–1070.

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.

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

Petrin, A., and K. Train (2004): "Omitted Product 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.

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.

Tauchen, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.

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.

Appendix

We review assumptions 1-8 from Newey (1985) in the context of our discrete choice setting. The observed data in this appendix is denoted generically as Y_i. $f(y|\theta)$ is the density function for the econometric model under consideration. $\tau = (\theta, \beta_\xi)$ is the full parameter vector. For matrix $A = [a_{ij}]$, $|A| = max_{i,j}|a_{ij}|$.

Assumption 1. *The observed data Y_i $i = 1, \ldots, N$ consists of random vectors which are independently distributed with common probability density function $h(y|\theta_0, \beta_{\xi 0})$ with respect to a measure space Y^*, where $\beta_{\xi 0}$ is an $L \times 1$ vector of parameters.*

Assumption 2. $\beta_\xi = \beta_{\xi 0}/\sqrt{N}$

β_ξ are the parameters that affect the correctness of the specification. In the case where ξ is observed this assumption allows one to derive the power properties of the test. When $\hat{\xi}$ is used the power properties will depend on the statistical properties of the error $\xi - \hat{\xi}$, which can be difficult to derive. For this reason, in the body of the paper we do not derive power properties, and thus do not require A2 to hold. Instead, we show the conditions under which the tests are consistent.

Assumption 3. *For all $\theta \in \Theta$ and almost all y in Y^*, $h(y|\theta, \beta_{\xi 0}) = f(y|\theta)$.*

This assumption states that the econometric model is correctly specified at the true value $\beta_{\xi 0}$.

Assumption 4. *The function $h(y|\tau)$ and the vector $m(y, \theta)$ are measurable functions of y for each (θ, β_ξ) in $B = \Theta \times \Gamma$, where Θ and Γ are compact subsets of R^R and R^L respectively and y is an element of a measure space Y^*. Also, for almost all $y \in Y^*$, $h(y|\tau)$ is twice continuously differentiable and*

$m(y,\theta)$ is once continuously differentiable in B and $(\theta_0', \beta_{\xi_0}')'$ is an element of the interior of B.

Assumption 5. *There exists measurable functions $a(y)$ and $b(y)$ such that for almost all y, $h(y|\tau) \le a(y)$ and $|lnh(y|\tau)|$, $|\frac{\partial lnh(y|\tau)}{\partial \tau}|^2$, $|\frac{\partial^2 lnh(y|\tau)}{\partial \tau \partial \tau'}|$, $|m(y,\theta)|^2$ and $|\frac{\partial m(y,\theta)}{\partial \theta}|$ are each less than $b(y)$ for all τ in B. It is also the case that $\int a(y)dy < +\infty$ and $\int b(y)a(y)dy < +\infty$, and that the set $\{y : h(y|\tau) > 0\}$ is independent of τ.*

Assumptions 4 and 5 are regularity and dominance conditions; as Newey (1985) notes, while not the weakest possible set of sufficient conditions, they are typically easy to verify.

Assumption 6. *If $\theta \neq \theta_0$, then $A = \{y|f(x|\theta) \neq f(y|\theta_0)\}$ satisfies $\int_A (f(x|\theta_0)dx > 0$.*

This is the usual identification assumption for θ_0.

Assumption 7. *For all $\theta \in \Theta$, $\int m(y,\theta) f(y|\theta)dy = 0$.*

This assumption formally states that that the moment condition(s) used for testing have mean zero in the population.

Assumption 8. *The matrix $V = E[(m_i(y,\theta_0)', s_i(\theta_0)')'(m_i(y,\theta_0)', s_i(\theta_0)')]$ is non-singular.*

This assumption rules out the possibility of linear dependencies among the moments.