

NBER WORKING PAPER SERIES

TESTS FOR PRICE ENDOGENEITY IN DIFFERENTIATED PRODUCT MODELS

Kyoo il Kim  
Amil Petrin

Working Paper 19011  
<http://www.nber.org/papers/w19011>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2013

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Kyoo il Kim and Amil Petrin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Tests for Price Endogeneity in Differentiated Product Models  
Kyoo il Kim and Amil Petrin  
NBER Working Paper No. 19011  
May 2013  
JEL No. C3,L0

**ABSTRACT**

We develop simple tests for endogenous prices arising from omitted demand factors in discrete choice models. Our approach only requires one to locate testing proxies that have some correlation with the omitted factors when prices are endogenous. We use the difference between prices and their predicted values given observed demand and supply factors. If prices are exogenous, these proxies should not explain demand given prices and other explanatory variables. We reject exogeneity if these proxies enter significantly in utility as additional explanatory variables. The tests are easy to implement as we show with several Monte Carlos and discuss for three recent demand applications.

Kyoo il Kim  
Department of Economics  
University of Minnesota  
4-129 Hanson Hall  
1925 4th Street South  
Minneapolis, MN 55455  
kyookim@umn.edu

Amil Petrin  
Department of Economics  
University of Minnesota  
4-101 Hanson Hall  
Minneapolis, MN 55455  
and NBER  
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 to 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 upwards. 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)).

The literature has responded to these findings in three distinct ways. The first is to integrate out the unobserved attributes. The challenge here is that one must specify the distribution of the demand errors conditional on observed characteristics and price. This is generally regarded as a difficult object to model.

The second has been to invert out the demand errors and then directly condition on them. Usually this is done by inverting the market share equations (Berry (1994), Berry et al. (1995), Gandhi, Kim, and Petrin (2011)). A drawback for these inversions is that they are computationally complex to program and time-consuming, and are also subject to propagated numerical errors from the inversion (see e.g. Dube, Fox, and Su (2012) and Knittel and Metaxoglou (2008)). One can also invert the demand error from the pricing equation but this requires stronger assumptions (Petrin and Train (2010), Kim and Petrin (2010)).

The final approach is to assert that there is no endogeneity problem conditional on the set of observed controls. For example, in some applications researchers observe household-level panels repeatedly choosing from the same choice sets over time. In this case one can include fixed effects directly in the estimation for goods whose physical characteristics are not changing over time. In some data sets many of the promotional activities are also known, like when the product is on display or promotion. Conditional on all of these factors some researchers then proceed as if there were no problem or that it is so small given the observed factors that it is not empirically relevant.<sup>1</sup>

In this paper we develop a class of asymptotic 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 (Petrin and Train (2010)). We develop control function type tests in the spirit of Smith and Blundell (1986), Rivers and Vuong

---

<sup>1</sup>See Chintagunta, Dube, and Goh (2005) for suggestive evidence that even data with this detail still suffers from omitted attributes.

(1988), and Villas-Boas and Winer (1999).<sup>2</sup>

A major advantage of our testing approaches is their simplicity. In the simplest format, all of the tests use least squares (or nonparametric regression) in an initial stage to obtain the proxies that enter the control functions. The proxies are calculated as the difference between product price and its predicted value given the relevant demand and/or supply factors that the econometrician observes. Our procedure then includes the controls directly in the maximization of the likelihood function and tests their significance, rejecting “exogenous prices” if they enter significantly. The controls can enter additively or interactions between price and the controls can be included to look for non-separability.

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 correctly identify the price endogeneity problem in almost every case.

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) and rejects price exogeneity.<sup>3</sup>

Our work is related to recent developments in nonparametric identification of the unobserved attribute in the discrete choice demand models (see Berry and Haile (2010) and Berry, Gandhi, and Haile (2011)). One could develop a test for price endogeneity by using this nonparametrically recovered unobserved attribute and combining it with more recent conditional mean tests or specification tests in general. There have been two main approaches in the literature. One approach is to use smoothed conditional mean functions (e.g. Härdle and Mammen (1993), Fan and Li (1996), Horowitz and Spokoiny (2001), Kitamura, Tripathi, and Ahn (2004) among others) and a second approach is to formulate a conditional mean restriction as a dense set of unconditional mean restrictions (e.g. Bierens (1990) and his other works, Stinchcombe and White (1998), Andrews (1997), Chen and Fan (1999), Li, Hsiao, and Zinn (2003), Song (2010) among others). These approaches may have better power properties and also be less sensitive to parametric assumptions. We leave them as future research and in this paper we focus on the classical tests based on parametric demand model settings that are most commonly used by practitioners, such as the random coefficient logit demand model.

The paper proceeds as follows. Section 2 describes differentiated products demand models and the endogeneity problem. Sections 3-5 describes the theory of the tests and their empirical

---

<sup>2</sup>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.

<sup>3</sup>In the separate work (Kim and Petrin (2010)) we effectively run a variant of this test for the automobile demand specification from BLP and the margarine demand specification from CDG.

implementation. Sections 6-7 includes the Monte Carlos and the empirical example. Then we conclude in Section 8. Technical details are added in Appendix.

## 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 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  $K \ll J$  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.

We analyze the specification problem that arises when the  $K$  factors used by the econometrician exclude relevant characteristics. We operate in the generalized random coefficients setting from Gandhi, Kim, and Petrin (2011) that allows for a non-separable error, with utility that consumer  $i$  derives from product  $j$  in market  $m$  is given as

$$U_{imj} = c_{ij} + \sum_{k=1}^K x_{mjk} \beta_{ik} - \alpha_i p_{mj} + \beta_{i\xi} \xi_{mj} + \sum_{k=1}^K \gamma_{ik} x_{mjk} \xi_{mj} + \gamma_{ip} p_{mj} \xi_{mj} + \epsilon_{imj}, \quad (1)$$

with  $x_{mjk}$  and  $p_{mj}$  being characteristics and price and the outside good 0 normalized to  $U_{im0} = \epsilon_{im0}$ .<sup>4</sup> In this setup coefficients are varying across individual observations according to (e.g.)

$$\begin{aligned} c_{ij} &= c_{j0} + \sum_{l=1}^L c_{jl} d_{il} + \sigma_c \omega_{ic} \\ \beta_{ik} &= \beta_{k0} + \sum_{l=1}^L \beta_{kl} d_{il} + \sigma_k \omega_{ik} \\ \alpha_i &= \alpha_0 + \sum_{l=1}^L \alpha_l d_{il} + \sigma_p \omega_{ip} \end{aligned}$$

where a vector of characteristics  $d_i$  ( $L \times 1$ ) can include demographics or other observed characteristics of  $i$  and  $\omega_i = (\omega_{ic}, \omega_{i1}, \dots, \omega_{iK}, \omega_{ip})$  denotes a vector of unobserved consumer tastes for the characteristics and the price (i.e., random shocks that generate the random coefficients), which follows a known distribution such as standard normals.<sup>5</sup> The idiosyncratic consumer-product specific shock  $\epsilon_{im} = (\epsilon_{im1}, \epsilon_{im2}, \dots, \epsilon_{imJ})'$  is independent across individuals  $i = 1, 2, \dots, N$  with joint distribution denoted by  $f_\epsilon(\cdot)$  that is known up to a finite vector of parameters.

$\xi_{mj}$  is a product-specific characteristic known to both consumers and producers in the market but unobserved to the econometrician. We isolate it from the other characteristics because the specification question asks whether  $\xi_{mj}$  enters utility for product  $j$  (i.e. whether  $\beta_{i\xi} \neq 0$ ,  $\gamma_{ik} \neq 0$ ,

<sup>4</sup>Gandhi, Kim, and Petrin (2011) show that in the automobile data from Berry, Levinsohn, and Pakes (1995) allowing for these interactions leads to an increase in price elasticities of 60% on average.

<sup>5</sup>The market index  $m$  can 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.

or  $\gamma_{ip} \neq 0$ ). For notational simplicity let  $\theta$  collect all the parameters in the model (1).<sup>6</sup> Letting  $(X, p)$  be the  $J \times (K + 1)$  matrix of the entire set of characteristics with arbitrary row  $(X'_j, p_j)$ ,  $\xi = (\xi_1, \dots, \xi_J)'$ , and  $P_{imj}(\theta)$  the probability of  $i$  choosing  $j$  conditional on  $(X, p, \xi, d)$ , and assuming that  $\epsilon$  and  $\omega$  are independent each other and from  $(X, p, \xi, d)$ , the choice probability  $P_{imj}(\theta)$  is given by

$$P_{imj}(\theta) = \int 1\{U_{imj} > U_{imk} \ \forall k \neq j | X, p, \xi, d_i, \theta\} dF_\epsilon(\epsilon) dF_\omega(\omega) \quad (2)$$

where  $F_\epsilon(\cdot)$  and  $F_\omega(\cdot)$  denote the distributions of  $\epsilon$  and  $\omega$ , respectively.

The error term in utility is given by

$$e_{imj} = \beta_i \xi \xi_{mj} + \sum_{k=1}^K \gamma_{ik} x_{mjk} \xi_{mj} + \gamma_{ip} p_{mj} \xi_{mj} + \epsilon_{imj}.$$

When price is in part determined by the omitted attribute, which one can think of as the product-market demand error, price is endogenous if  $\beta_i \xi \neq 0$ . For example, if sellers charge higher prices when their products have more desirable omitted characteristics, the price will be positively correlated with the demand error, biasing price elasticities upwards. When the interaction terms  $\gamma_{ik} \neq 0$  or  $\gamma_{ip} \neq 0$ , the econometric problem becomes more complicated, as there are two or more terms in the error that are not independent of price.<sup>7</sup>

There have been two main approaches in estimation to deal with an unobserved demand attribute that is correlated with price. We illustrate them with the conditional likelihood function for this problem, given as

$$\log L_N(Y|X, p, \xi, d, \theta) = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J Y_{imj} \log P_{imj}(\theta),$$

where  $N = \sum_{m=1}^M N_m$  and  $Y_{imj}$  is an indicator variable that equals one if  $i$  is observed to choose  $j$  at a market  $m$ .<sup>8</sup> The first approach is to integrate out the unobserved attribute. This requires one to specify the distribution of the demand errors conditional on observed characteristics and price -  $F_{\xi|X,p}(\xi|X, p)$  - and then integrate them out in the calculation of the choice probability given in (2). The challenge is in the specification of  $F_{\xi|X,p}(\xi|X, p)$ . It requires that one first know  $F_{p|X,\xi}(p|X, \xi)$ , the distribution of the vector of market prices conditional on all observed and unobserved product characteristics, and then that one be able to recover  $F_{\xi|X,p}(\xi|X, p)$  from  $F_{p|X,\xi}(p|X, \xi)$ .

The second approach is to condition directly on the unobserved vector  $\xi$  by inverting it out of the demand equations following Berry, Levinsohn and Pakes (1995) (or the generalization of it in Gandhi, Kim, and Petrin (2011) when  $\gamma_{ik} \neq 0$  or  $\gamma_{ip} \neq 0$ ). While these approaches allow the

---

<sup>6</sup>For fixed parameter models (the coefficients do not vary across  $i$ ), we have  $\theta = (c', \beta', \alpha, \beta_\xi, \gamma', \gamma_p)'$  and for heterogeneous coefficients models in addition to these (mean) parameters,  $\theta$  also includes the standard deviations ( $\sigma = (\sigma_c, \sigma_1, \dots, \sigma_K, \sigma_p)$ ) of the random coefficients and other coefficients on demographics or other observed characteristics of households.

<sup>7</sup>Here we focus on the endogeneity of price but other observable attributes can also be correlated with the unobserved demand factor. Indeed the rejection of the price exogeneity in our test can arise due to the endogeneity of other attributes. We view all of these as an evidence for the endogeneity or the omitted attribute.

<sup>8</sup>When only the aggregate (market-level) data is available, we replace  $Y_{imj}$  with  $s_{mj}$ , the market share of product  $j$  at a market  $m$  and let  $P_{mj}(\theta)$  be the choice probability.

researcher to avoid specifying  $F_{\xi|X,p}(\xi|X,p)$ , they are computationally complex to program and time-consuming, and are also subject to propagated numerical errors from the inversion, which can lead to non-convergence in the estimation step, as noted recently in the literature (e.g. Dube, Fox, and Su (2012) and Knittel and Metaxoglou (2008)). We now turn to explaining our computationally fast and simple tests which inform the researcher if he/she needs to worry about the presence of an unobserved demand attribute.

### 3 Using Proxies to Test if Unobserved Attributes Enter Utility

In this section we develop test statistics - Wald and LM tests - that use proxies for the unobserved attributes to test for their presence in utility. Our testing philosophy is easiest to understand if we start by treating the unobserved attribute as if it were observed but omitted from the specification. For expositional convenience we assume that the taste for the omitted attribute is common across individuals, so  $\beta_{i\xi} = \beta_\xi$ ,  $\gamma_i = \gamma$ , and  $\gamma_{ip} = \gamma_p$  for all  $i$ .<sup>9</sup>

We refer to our test as the *price endogeneity* test because the null hypothesis presumes  $\xi$  does not enter utility either in an additive and non-additive utility manner:

$$H_0 : \beta_\xi = 0, \gamma = 0, \text{ and } \gamma_p = 0. \quad (3)$$

We then deem prices as exogenous if this null hypothesis is true.<sup>10</sup> Both the Wald and the LM formulation of the test of either hypothesis would be consistent under the alternative. In the former case the terms related to the omitted attribute would be statistically significant in the unconstrained specification. In the latter case where the constrained model is estimated without the omitted attribute terms, the residual  $(Y_{imj} - P_{imj}(\theta))$  from this setup would be significantly correlated with these omitted terms.<sup>11</sup>

Since the omitted attribute is not observed, we propose *feasible* tests that are analogs to the aforementioned tests but are instead based on using a proxy for the unobserved attribute. Letting  $\tilde{\xi}_{mj}$  denote the proxy for  $\xi_{mj}$ , consistency of the tests only requires that - under the alternative - there is some dependence between  $\xi_{mj}$  and  $\tilde{\xi}_{mj'}$  for at least one  $j'$ , typically for  $j' = j$ .

We exploit the source of the endogeneity problem to develop a proxy that should satisfy this condition. Specifically, letting  $Z_m = (X_m, Z_{2m})$ , where  $Z_{2m}$  denotes the supply-side factors observable to econometrician, equilibrium prices  $p_{mj}$  are in part determined by the demand shocks  $\xi_m$  conditional on  $Z_m$ :

$$Cov(p_{mj}, \xi_{mj'} | Z_m) \neq 0 \quad (4)$$

---

<sup>9</sup>Extensions allowing for the heterogeneous coefficients on the omitted attribute and/or the interaction terms are straightforward using more parameters.

<sup>10</sup>If one only wanted to test for unobserved attributes that entered utility linearly then this null would become  $H_0^1 : \beta_\xi = 0$ .

<sup>11</sup>For the test of (3) with *observed*  $\xi$  one can use classical asymptotic tests such as Wald, LM, conditional mean tests in nonlinear models, see e.g. Newey (1985) and Newey and McFadden (1994).

where the covariance is typically largest for  $j' = j$ . For our proxies we use the mean-squared projection residuals of  $p_{mj}$  on  $Z_m$ :

$$\tilde{\xi}_{mj} = p_{mj} - E[p_{mj}|Z_m], \quad (5)$$

as  $\tilde{\xi}_m = (\tilde{\xi}_{m1}, \dots, \tilde{\xi}_{mJ})'$  from (5) will not generally be independent of  $\xi_{mj}$ .<sup>12</sup> In Section 5.1, we provide a more technical discussion of the conditions under which the tests are consistent.

Our approach then replaces the unobserved  $\xi_{mj}$  in utility with some chosen functions of  $\tilde{\xi}_m$ , denoted as  $v_j(\tilde{\xi}_m)$ , where  $v_j(\cdot)$  can be a vector function and  $v(\cdot) = (v_1(\cdot), \dots, v_J(\cdot))'$ . In practice one can let  $v_j(\tilde{\xi}_m) = \tilde{\xi}_{mj}$ , or one can include  $\tilde{\xi}_{mj}$  as the leading term when  $v_j(\cdot)$  is a vector. We approximate  $\tilde{\xi}_{mj}$  with  $\hat{\xi}_{mj}$ , the difference between  $p_{mj}$  and the estimated  $\hat{E}[p_{mj}|Z_m]$ :  $\hat{\xi}_{mj} = p_{mj} - \hat{E}[p_{mj}|Z_m]$ ,  $j = 1, \dots, J$ . We then replace  $\xi_{mj}$  with a function of  $\hat{\xi}_m$ ,  $v_j(\hat{\xi}_m)$  (in the simplest case we choose  $v_j(\hat{\xi}_m) = \hat{\xi}_{mj}$ ). We discuss estimation of  $E[p_{mj}|Z_m]$  for a variety of data generating processes in Section 4.

### 3.1 The Wald test

The Wald test includes the controls directly in the estimated specification. Letting  $\lambda$  denote the parameter on the controls, the utility is specified for estimation as

$$\tilde{U}_{imj} = c_{ij} + \sum_{k=1}^K x_{mjk} \beta_{ik} - \alpha_i p_{mj} + \lambda'_j v_j(\tilde{\xi}_m) + \sum_{k=1}^K \tilde{\gamma}'_k v_j(\tilde{\xi}_m) x_{mjk} + \tilde{\gamma}'_p v_j(\tilde{\xi}_m) p_{mj} + \epsilon_{imj}, \quad (6)$$

with  $v_j(\tilde{\xi}_m)$  as the control functions. Under the null hypothesis of exogenous prices the coefficients  $\lambda = (\lambda'_1, \dots, \lambda'_J)'$ ,<sup>13</sup>  $\tilde{\gamma} = (\tilde{\gamma}'_1, \dots, \tilde{\gamma}'_K)'$ , and  $\tilde{\gamma}_p$  are equal to zero because  $\tilde{\xi}_m$  does not affect the demand given  $(x_m, p_m, d_i)$ . Therefore we test for the price endogeneity due to the unobserved attribute by testing the null hypotheses

$$\tilde{H}_0 : \lambda = 0, \tilde{\gamma} = 0, \text{ and } \tilde{\gamma}_p = 0. \quad (7)$$

This is the feasible version of the test in (3).

For the formulation of the Wald statistic let  $\tilde{\theta} = (\vartheta', \lambda', \tilde{\gamma}', \tilde{\gamma}'_p)'$ , where  $\vartheta$  denotes all parameters excluding  $(\lambda', \tilde{\gamma}', \tilde{\gamma}'_p)'$ , and write the sample ML objective function that is used to estimate  $\tilde{\theta}$  as

$$\log L_N(\tilde{\xi}_m, \tilde{\theta}) = \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J Y_{imj} \log P_{imj}(v(\tilde{\xi}_m), \tilde{\theta})/N.$$

Then we obtain the unconstrained parameter estimates  $\hat{\theta}_U$  that solve the sample analog equations

$$\frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_U)}{\partial \tilde{\theta}} \equiv \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J Y_{imj} \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\theta}_U)}{\partial \tilde{\theta}} = 0$$

<sup>12</sup>If the dependence of  $p_{mj}$  on  $\xi_m$  conditional on  $Z_m$  is not reflected in their conditional covariance - if they are related by higher moments - then the test using the simple proxy (control)  $\tilde{\xi}_m$  may not reject the null even when the price is endogenous. If one suspected this were there are alternative ways of obtaining the controls robust to this situation (e.g., Matzkin (2003), Altonji and Matzkin (2005), and Imbens and Newey (2003)).

<sup>13</sup>One can restrict  $\lambda_1 = \dots = \lambda_J$  depending on applications.

where the conditional choice probability  $P_{imj}(v(\hat{\xi}_m), \tilde{\theta})$  is obtained based on the specification of utility in (6) but with  $v(\hat{\xi}_m)$  replacing  $v(\tilde{\xi}_m)$ . The Wald tests ask whether the coefficients  $(\hat{\lambda}', \hat{\gamma}', \hat{\gamma}'_p)'$  from  $\hat{\theta}_U$  are statistically different from zero. The test statistic is

$$\tilde{T}_{M,N} = N(\hat{\lambda}', \hat{\gamma}', \hat{\gamma}'_p) \tilde{V}_{M,N}^{-1} (\hat{\lambda}', \hat{\gamma}', \hat{\gamma}'_p)'$$

where  $\tilde{V}_{M,N}$  is a consistent estimator of the asymptotic variance matrix of the ML estimator  $(\hat{\lambda}', \hat{\gamma}', \hat{\gamma}'_p)'$  that accounts for the estimated controls.  $\tilde{T}_{M,N}$  follows the asymptotic  $\chi^2$  distribution with the degree of freedom equal to the dimension of  $(\lambda', \gamma', \gamma'_p)'$ . We derive the asymptotic distribution of the test statistics in Section 5. For definition of  $\tilde{V}_{M,N}$  see Theorem 5.1.

The Wald approach is attractive for three reasons. The test statistic is easy to compute, requiring no additional calculation beyond standard likelihood function output, and significance of the coefficients is equivalent to rejecting price exogeneity. The resulting estimates may be suggestive of the direction and magnitude of the bias arising from the price endogeneity. Finally, the point estimates can be used as starting values for the more complicated correction approaches. Good starting values can be helpful when the objective function is non-linear in parameters, there are a large number of parameters, or there are possibly multiple local maxima, which are often characteristics of these alternative approaches.

### 3.2 The LM test: Constrained Approach

The LM test imposes  $\lambda = 0$ ,  $\tilde{\gamma} = 0$ , and  $\tilde{\gamma}_p = 0$  during estimation and then tests to see if the score function, which is the derivative of the likelihood with respect to the parameter  $\tilde{\theta}$  evaluated at the constrained estimates, is large in absolute value. From a practical standpoint the LM test is useful because it does estimation in the constrained case, which can reduce the computational cost when the number of parameters being tested is large. We start with the test for the price endogeneity and then develop the test for non-additivity.

Again let  $\tilde{\theta} = (\vartheta', \lambda', \tilde{\gamma}', \tilde{\gamma}'_p)'$  and let  $D = \dim((\lambda', \tilde{\gamma}', \tilde{\gamma}'_p))$ . Then let  $\tilde{\theta}_R$  be the constrained parameter  $\tilde{\theta}_R = (\vartheta', 0'_D)'$  where  $0_D$  denotes a  $D \times 1$  vector of zeros. The maximum likelihood estimator in the constrained case uses the sample analogs to the  $D$  population moments

$$E\left[\sum_{j=0}^J Y_{imj} \frac{\partial \log P_{imj}(v(\tilde{\xi}_m), \tilde{\theta}_R)}{\partial \vartheta}\right] = 0.$$

The constrained parameter estimates  $\hat{\tilde{\theta}}_R = (\hat{\vartheta}', 0'_D)'$  solve the sample analog equations

$$\frac{\partial \log L_N(\hat{\xi}_m, \hat{\tilde{\theta}}_R)}{\partial \vartheta} \equiv \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J Y_{imj} \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\tilde{\theta}}_R)}{\partial \vartheta} = 0. \quad (8)$$

The LM test is based on the average value of the derivative of the likelihood with respect to  $\tilde{\theta}$  evaluated at the constrained parameter estimates:

$$\left. \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta})}{\partial \tilde{\theta}} \right|_{\tilde{\theta}=\hat{\tilde{\theta}}_R} \equiv \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J Y_{imj} \left. \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \tilde{\theta})}{\partial \tilde{\theta}} \right|_{\tilde{\theta}=\hat{\tilde{\theta}}_R}.$$

With abuse of notation, we often write  $\frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_R)}{\partial \theta} = \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta})}{\partial \theta} \Big|_{\tilde{\theta}=\hat{\theta}_R}$  and also others similarly.

Now construct a nonrandom matrix  $H$  (with the number of rows equal to the dimension of  $(X', \tilde{\gamma}', \tilde{\gamma}'_p)'$ ) such that  $H\tilde{\theta} = 0$  becomes the null hypothesis of the price exogeneity (7). Then by construction we have  $H\hat{\theta}_R = 0$ . Let  $\hat{\Gamma}(\tilde{\theta}) = -\frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta})}{\partial \theta \partial \theta'}$ . Then the LM test statistic for price endogeneity is given by

$$\widetilde{LM}_{M,N} = N \frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_R)}{\partial \theta'} \hat{\Gamma}(\hat{\theta}_R)^{-1} H' \tilde{V}_{M,N}^{-1}(\hat{\theta}_R) H \hat{\Gamma}(\hat{\theta}_R)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_R)}{\partial \theta}$$

where  $\tilde{V}_{M,N}(\hat{\theta}_R)$  is a consistent estimator of the asymptotic variance matrix of  $\sqrt{N} H \hat{\Gamma}(\hat{\theta}_R)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_R)}{\partial \theta}$ .

Under the null hypothesis of (7),  $\widetilde{LM}_{M,N}$  follows the same asymptotic  $\chi^2$  distribution with the corresponding Wald test statistic. See Theorem 5.2 below for this result and also for the construction of  $\tilde{V}_{M,N}(\hat{\theta}_R)$ .

## 4 Proxy Estimation Across Data Generating Processes

The purpose of estimating  $E[p_{mj}|Z_m]$  is to recover the proxy  $\hat{\xi}_{mj} = p_{mj} - \hat{E}[p_{mj}|Z_m]$ . This step can be done by least squares using polynomials of  $Z_m$  as regressors or by nonparametric series estimation (e.g. Newey (1997)). For the Wald test to have power  $v_j(\hat{\xi}_m)$  cannot be perfectly collinear with  $X_{mj}$  and  $p_{mj}$ . For the LM test to have power the score associated with the coefficients on  $v_j(\hat{\xi}_m)$  should not be perfectly collinear with the scores associated with the coefficients on  $X_{mj}$  and  $p_{mj}$ . A sufficient condition is that a variable excluded from  $X_{mj}$  in the instruments set is correlated with prices (e.g., either characteristics of other products if they are assumed to be exogenous or other observed supply-side factors).

The power of the test is increasing in  $Cov(\xi_{mj}, v_j(\hat{\xi}_m))$ , which is determined by the variation that is used to estimate  $E[p_{mj}|Z_m]$  and the restrictions placed on  $E[p_{mj}|Z_m]$  during estimation. This variation comes 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  $E[p_{mj}|Z_m]$  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  $E[p_{mj}|Z_m]$  for each product using the cross-market variation (allowing the

coefficients of each function to differ for each type of product).  $Z_m$  includes  $X_{mj}$  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  $E[p_{mj}|Z_m]$  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  $E[p_{mj}|Z_m]$  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, 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  $Cov(p_{mj}, \xi_{mj}|Z_m) > 0$ , a large residual is suggestive of more marketing-mix activities.

## 5 Asymptotic Distributions of Test Statistics

In this section we derive the asymptotic distributions of the Wald test and the LM test statistics. Our asymptotic experiment is mainly in the number of markets  $M$  as  $M \rightarrow \infty$  while  $N_m$  (the number of individuals in each market) can be finite or tend to infinity. In both cases note, however, that  $N = \sum_{m=1}^M N_m$  also tends to infinity as  $M \rightarrow \infty$ . We derive the asymptotic distribution of

the Wald test statistic accounting for the estimated controls.<sup>14</sup> The controls  $\hat{\xi}_m$  are obtained from regressions of  $P_m$  on  $Z_m$  as

$$\tilde{\xi}_{mj} = p_{mj} - \mathbf{Z}'_m \pi_j, \quad j = 1, \dots, J$$

where  $\mathbf{Z}_m = Z_m$  or  $\mathbf{Z}_m$  can also include higher order terms of  $Z_m$ . We can also extend this (and results below) to nonparametric series estimation but the estimated standard errors under the parametric regression will be equivalent to those that would be obtained under the nonparametric regression given approximation anyway (e.g. Newey (1997)), so we present our results in the parametric flexible regression context.

Note that  $\sqrt{M}(\hat{\pi}_j - \pi_{j0}) = (\frac{1}{M} \sum_{m=1}^M \mathbf{Z}_m \mathbf{Z}'_m)^{-1} \frac{1}{\sqrt{M}} \sum_{m=1}^M \mathbf{Z}'_m \tilde{\xi}_{mj}$  and we let  $\sqrt{M}(\hat{\pi}_j - \pi_{j0}) \rightarrow_d \varpi_j$  where  $(\varpi_1, \dots, \varpi_J)$  follow a joint normal distribution. We write  $\tilde{\xi}_{mj}(\pi) = p_{mj} - \mathbf{Z}'_m \pi_j$ ,  $\hat{\xi}_{mj} = \tilde{\xi}_{mj}(\hat{\pi})$ , and  $\tilde{\xi}_{mj} = \tilde{\xi}_{mj}(\pi_0)$  for  $j = 1, \dots, J$ .

The regularity conditions below that validate the asymptotic distributions and the consistency of the variance matrix estimators are fairly standard (see Newey and McFadden (1994) for example). We collect a set of sufficient conditions that render the asymptotic  $\chi^2$  distribution of  $\tilde{T}_{M,N}$  here and also that of  $\tilde{LM}_{M,N}$ . We let  $\pi = (\pi'_1, \dots, \pi'_J)'$  and let  $\|A\| = \sqrt{\text{tr}(A'A)}$  for a matrix or a vector  $A$ . Also let  $\Sigma_{\mathbf{Z}\mathbf{Z}} = \text{plim}_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \mathbf{Z}_m \mathbf{Z}'_m$  and  $\Sigma_{\tilde{\xi}^2 \mathbf{Z}\mathbf{Z}} = \text{plim}_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \tilde{\xi}_{mj}^2 \mathbf{Z}_m \mathbf{Z}'_m$ .

**Assumption 5.1** (i)  $\{p_m, Z_m\}_{m \leq M}$  are independently distributed across  $m$  and  $\{Y_{im1}, \dots, Y_{imJ}, d_i\}_{i \leq N_m}^{m \leq M}$  are independently and identically distributed across  $i$  and are independently distributed across  $m$ ; (ii)  $E[\|\tilde{\xi}_m\|^4] < \infty$  and  $E[\|\mathbf{Z}_m\|^4] < \infty$  for all  $m$ ; (iii)  $\Sigma_{\mathbf{Z}\mathbf{Z}}$  is nonsingular.

Then under Assumption 5.1  $\sqrt{M}(\hat{\pi}_j - \pi_{j0}) \rightarrow_d \varpi_j \sim \mathcal{N}(0, \Sigma_{\mathbf{Z}\mathbf{Z}}^{-1} \Sigma_{\tilde{\xi}^2 \mathbf{Z}\mathbf{Z}} \Sigma_{\mathbf{Z}\mathbf{Z}}^{-1})$  for all  $j$  by the Lindeberg-Feller CLT. We let  $\tilde{\theta}_0$  denote the true value of  $\tilde{\theta}$ .

**Assumption 5.2** (i)(a)  $\tilde{\theta}_U \rightarrow_p \tilde{\theta}_0$ , (b)  $\tilde{\theta}_R \rightarrow_p \tilde{\theta}_0$  under the null (7); (ii)  $\tilde{\theta}_0$  is in the interior of  $\tilde{\Theta}$ ; (iii)  $P_{imj}(v(\tilde{\xi}_m), \tilde{\theta})$  is twice continuously differentiable in  $\tilde{\theta}$  on some neighborhood  $\tilde{\Theta}_0$  of  $\tilde{\theta}_0$  ( $\tilde{\Theta}_0 \subset \tilde{\Theta}$ ); (iv)  $P_{imj}(v(\tilde{\xi}_m), \tilde{\theta})$  is Lipschitz in  $\tilde{\xi}_m$  for all  $j$ ; (v)  $\frac{\partial^2 \log P_{imj}(v(\tilde{\xi}_m(\pi)), \tilde{\theta})}{\partial \theta \partial \theta'}$  is continuous in  $(\tilde{\theta}, \pi)$  for all  $w \in \mathcal{W}$ ,  $\mathcal{W}$  denotes the support of  $\tilde{W}_{mi} = (X'_m, p'_m, \tilde{\xi}'_m, d_i)'$  and  $E[\frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \theta \partial \theta'}]$  is nonsingular for all  $M$  large enough; (vi)  $E[\sup_{\tilde{\theta} \in \tilde{\Theta}_0, \pi \in \Pi_0} \|\frac{\partial \log P_{imj}(v(\tilde{\xi}_m(\pi)), \tilde{\theta})}{\partial \theta \partial \theta'}\|] < \infty$  for all  $j$  and  $m$  where  $\Pi_0$  denotes some neighborhood of  $\pi_0$ ; (vii)  $E[\sup_{\tilde{\theta} \in \tilde{\Theta}_0, \pi \in \Pi_0} \|\frac{\partial \log P_{imj}(v(\tilde{\xi}_m(\pi)), \tilde{\theta})}{\partial \theta}\|^4] < \infty$  for all  $j$  and  $m$ ; (viii)  $\frac{\partial^2 \log P_{imj}(v(\tilde{\xi}_m(\pi)), \tilde{\theta})}{\partial \theta \partial v_{j'}(\tilde{\xi}_m)}$  is continuous in  $(\tilde{\theta}, \pi)$  for all  $w \in \mathcal{W}$  and  $E[\sup_{\tilde{\theta} \in \tilde{\Theta}_0, \pi \in \Pi_0} \|\frac{\partial^2 \log P_{imj}(v(\tilde{\xi}_m(\pi)), \tilde{\theta})}{\partial \theta \partial v_{j'}(\tilde{\xi}_m)}\|^2] < \infty$  for all  $j, j'$ , and  $m$ .

To obtain the asymptotic distribution of the Wald test statistic we first derive the asymptotic distribution of the unconstrained ML estimator that uses the estimated controls.

<sup>14</sup>Our asymptotic distributions here do not include simulation errors to approximate (e.g.) the distribution of random coefficients. If one uses simulations to approximate the distribution of random coefficients, one can add an additional variance term that arises from the simulation as in BLP.

**Lemma 5.1** *Suppose Assumptions 5.1 and 5.2 hold.*

Then  $\sqrt{M}(\widehat{\theta}_U - \widetilde{\theta}_0) \rightarrow_d \mathcal{N}(0, \Gamma_0^{-1}(\Gamma_0 \cdot \lim_{M \rightarrow \infty} \frac{M}{N} + V_1)\Gamma_0^{-1})$  where  $\Gamma_0 = \lim_{M \rightarrow \infty} E[-\frac{\partial \log L_N(\widehat{\xi}_m, \widetilde{\theta}_0)}{\partial \widetilde{\theta} \partial \widetilde{\theta}'}]$ ,  $V_1 = \sum_{j,k} \Lambda_{0j} Cov(\varpi_j, \varpi_k) \Lambda_{0k}'$ , and  $\Lambda_{0j} = \lim_{M \rightarrow \infty} E[\Lambda_j^{N,M}(\widetilde{\theta}_0, \widetilde{\xi}_m)]$  with  $\Lambda_j^{N,M}(\widetilde{\theta}, \widetilde{\xi}_m) \equiv \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j''=0}^J Y_{imj''} \sum_{j'=1}^J \frac{\partial^2 \log P_{imj''}(v(\widetilde{\xi}_m), \widetilde{\theta})}{\partial \widetilde{\theta} \partial v_{j'}(\widetilde{\xi}_m)} \frac{\partial v_{j'}(\widetilde{\xi}_m)}{\partial \widetilde{\xi}_{mj}} \frac{\partial \widetilde{\xi}_{mj}}{\partial \pi_j} / N$ .

See Appendix A.1 for the proof.

To obtain a consistent estimator of the variance matrix, we use  $\widehat{\Gamma} = -\frac{\partial \log L_N(\widehat{\xi}_m, \widehat{\theta}_U)}{\partial \widehat{\theta} \partial \widehat{\theta}'}$  for  $\Gamma_0$  and  $\widehat{V}_1 = \sum_{j,k} \Lambda_j^{N,M}(\widehat{\theta}_U, \widehat{\xi}_m) \widehat{Cov}(\varpi_j, \varpi_k) \Lambda_k^{N,M}(\widehat{\theta}_U, \widehat{\xi}_m)'$  for  $V_1$  where  $\widehat{Cov}(\varpi_j, \varpi_k)$  is obtained from

$$\widehat{Cov}(\varpi_j, \varpi_k) = \left( \sum_{m=1}^M \mathbf{Z}_m \mathbf{Z}_m' / M \right)^{-1} \left( \sum_{m=1}^M \widehat{\xi}_{mj} \widehat{\xi}_{mk} \mathbf{Z}_m \mathbf{Z}_m' / M \right) \left( \sum_{m=1}^M \mathbf{Z}_m \mathbf{Z}_m' / M \right)^{-1}.$$

When one assumes homoskedasticity across  $m$  and zero correlations of  $\widetilde{\xi}_{mj}$  across products  $j \neq k$ ,  $V_1$  can simplify to  $V_1 = \sum_{j=1}^J \Lambda_{0j} (\Sigma \mathbf{Z} \mathbf{Z}')^{-1} E[\widetilde{\xi}_{mj}^2] \Lambda_{0j}'$ .

Next we derive the asymptotic distribution of the Wald test statistic.

**Theorem 5.1** *Suppose Assumptions 5.1 and 5.2 (i)(a) and (ii)-(viii) hold. Then for the Wald statistic we have  $\widetilde{T}_{M,N} = N(\widehat{\lambda}', \widehat{\gamma}', \widehat{\gamma}_p') \widetilde{V}_{M,N}^{-1}(\widehat{\lambda}', \widehat{\gamma}', \widehat{\gamma}_p')$   $\rightarrow_d \chi^2(\dim((\lambda', \gamma', \gamma_p')))$  with  $\widetilde{V}_{M,N} = H \widehat{\Gamma}^{-1}(\widehat{\Gamma} + \widehat{V}_1 \frac{N}{M}) \widehat{\Gamma}^{-1} H'$  where*

$$\begin{aligned} \widehat{\Gamma} &= -\frac{\partial \log L_N(\widehat{\xi}_m, \widehat{\theta}_U)}{\partial \widehat{\theta} \partial \widehat{\theta}'}, \\ \widehat{V}_1(\widehat{\theta}) &= \sum_{j,k} \Lambda_j^{N,M}(\widehat{\theta}, \widehat{\xi}_m) \widehat{Cov}(\varpi_j, \varpi_k) \Lambda_k^{N,M}(\widehat{\theta}, \widehat{\xi}_m)', \quad \widehat{V}_1 = \widehat{V}_1(\widehat{\theta}_U), \\ \Lambda_j^{N,M}(\widetilde{\theta}, \widetilde{\xi}_m) &= \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j''=0}^J Y_{imj''} \sum_{j'=1}^J \frac{\partial^2 \log P_{imj''}(v(\widetilde{\xi}_m), \widetilde{\theta})}{\partial \widetilde{\theta} \partial v_{j'}(\widetilde{\xi}_m)} \frac{\partial v_{j'}(\widetilde{\xi}_m)}{\partial \widetilde{\xi}_{mj}} \frac{\partial \widetilde{\xi}_{mj}}{\partial \pi_j}, \quad \text{and} \\ \widehat{Cov}(\varpi_j, \varpi_k) &= \left( \sum_{m=1}^M \mathbf{Z}_m \mathbf{Z}_m' / M \right)^{-1} \left( \sum_{m=1}^M \widehat{\xi}_{mj} \widehat{\xi}_{mk} \mathbf{Z}_m \mathbf{Z}_m' / M \right) \left( \sum_{m=1}^M \mathbf{Z}_m \mathbf{Z}_m' / M \right)^{-1}. \end{aligned}$$

**Proof.** Note that we use the nonrandom matrix  $H$  such that  $H \widetilde{\theta} = 0$  represents the null hypothesis of the price exogeneity (7). Then with the variance-covariance matrix estimator of  $\sqrt{N} H \widehat{\theta}_U$  as  $\widetilde{V}_{M,N} = H \widehat{\Gamma}^{-1}(\widehat{\Gamma} + \widehat{V}_1 \frac{N}{M}) \widehat{\Gamma}^{-1} H'$ , we obtain

$$\widetilde{T}_{M,N} = N(\widehat{\lambda}', \widehat{\gamma}', \widehat{\gamma}_p') \widetilde{V}_{M,N}^{-1}(\widehat{\lambda}', \widehat{\gamma}', \widehat{\gamma}_p') = M \cdot (H \widehat{\theta}_U)' \left\{ H \widehat{\Gamma}^{-1} \left( \widehat{\Gamma} \frac{M}{N} + \widehat{V}_1 \right) \widehat{\Gamma}^{-1} H' \right\}^{-1} H \widehat{\theta}_U.$$

Therefore, by the continuous mapping theorem,  $\widetilde{T}_{M,N} \rightarrow_d \chi^2$  with the degree of freedom equal to the dimension of  $(\lambda', \gamma', \gamma_p)'$  (i.e. the number of rows in  $H$ ) because Lemma 5.1,  $\widehat{\Gamma} \rightarrow_p \Gamma_0$ , and  $\widehat{V}_1 \rightarrow_p V_1$  implies  $\mathcal{Z} \equiv \sqrt{M} \left\{ H \widehat{\Gamma}^{-1} \left( \widehat{\Gamma} \frac{M}{N} + \widehat{V}_1 \right) \widehat{\Gamma}^{-1} H' \right\}^{-1/2} H \widehat{\theta}_U \rightarrow_d \mathcal{N}(0, I)$  and  $\widetilde{T}_{M,N}$  is the quadratic function of  $\mathcal{Z}$ . ■

Next we show that the LM test statistic has the same asymptotic distribution of the corresponding Wald test statistic.

**Theorem 5.2** Suppose Assumptions 5.1 and 5.2(i)(b) and (ii)-(viii) hold. Then for the LM statistic we have

$$\widetilde{LM}_{M,N} = N \frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_R)}{\partial \hat{\theta}'} \widehat{\Gamma}(\hat{\theta}_R)^{-1} H' \widetilde{V}_{M,N}^{-1}(\hat{\theta}_R) H \widehat{\Gamma}(\hat{\theta}_R)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_R)}{\partial \hat{\theta}} \rightarrow_d \chi^2(\dim((\lambda', \tilde{\gamma}', \tilde{\gamma}_p)'))$$

where  $\widetilde{V}_{M,N}(\hat{\theta}_R) = H \widehat{\Gamma}(\hat{\theta}_R)^{-1} (\widehat{\Gamma}(\hat{\theta}_R) + \widehat{V}_1(\hat{\theta}_R) \frac{N}{M}) \widehat{\Gamma}(\hat{\theta}_R)^{-1} H'$ .

See Appendix A.2 for the proof.

## 5.1 Test Consistency

Here we further discuss the test consistency of price endogeneity. The Wald test is consistent as long as  $\text{plim}_{M \rightarrow \infty} (\hat{\lambda}', \hat{\tilde{\gamma}}', \hat{\tilde{\gamma}}_p)' \neq 0$  under the alternative against (7) i.e.  $\text{plim}_{M \rightarrow \infty} \widetilde{T}_{M,N} = \infty$  under the alternative. Therefore the Wald test is consistent under the alternative as long as the unobserved attribute that is correlated with price enters utility and the proxy is correlated with the unobserved attribute.

The LM test is more involved. To understand the consistency of the LM test, we first need to reformulate the LM statistic as below. Note that  $\sum_{j=0}^J P_{imj}(\cdot, \tilde{\theta}) = 1$  implies  $\sum_{j=0}^J P_{imj}(\cdot, \tilde{\theta}) \frac{\partial \log P_{imj}(\cdot, \tilde{\theta})}{\partial \tilde{\theta}} = 0$ . We therefore can write  $\frac{\partial \log L_N(\cdot, \tilde{\theta})}{\partial \tilde{\theta}} = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J (Y_{imj} - P_{imj}(\cdot, \tilde{\theta})) \frac{\partial \log P_{imj}(\cdot, \tilde{\theta})}{\partial \tilde{\theta}}$ . Define

$$u_{imj}(\tilde{\theta}) = Y_{imj} - P_{imj}(\cdot, \tilde{\theta})$$

and also let  $u_{imj}(\hat{\theta}_R) = Y_{imj} - P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)$  with abuse of notation. Then the LM test statistic can be written as

$$\begin{aligned} \widetilde{LM}_{M,N} = & N \left( \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)}{\partial \hat{\theta}'} u_{imj}(\hat{\theta}_R) \right) \widehat{\Gamma}(\hat{\theta}_R)^{-1} H' \widetilde{V}_{M,N}^{-1}(\hat{\theta}_R) H \widehat{\Gamma}(\hat{\theta}_R)^{-1} \\ & \times \left( \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J u_{imj}(\hat{\theta}_R) \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)}{\partial \hat{\theta}} \right). \end{aligned}$$

Recall that  $\tilde{\theta}_0$  denotes the true value of  $\tilde{\theta}$ . Then  $\frac{\partial \log P_{imj}(v(\tilde{\xi}_m), \tilde{\theta}_0)}{\partial \tilde{\theta}}$  must be uncorrelated with  $u_{imj}(\tilde{\theta}_0)$  because  $\tilde{\theta}_0$  solves the moment condition

$$0 = E \left[ \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} \right] = E \left[ \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j=0}^J u_{imj}(\tilde{\theta}_0) \frac{\partial \log P_{imj}(v(\tilde{\xi}_m), \tilde{\theta}_0)}{\partial \tilde{\theta}} \right] \quad (9)$$

and this is not equal to zero for any  $\tilde{\theta} \neq \tilde{\theta}_0$ . Therefore the LM statistic looks at the covariance between  $\frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)}{\partial \hat{\theta}}$  and  $u_{imj}(\hat{\theta}_R)$  to detect the violation of this moment condition, which plays the role of an evidence against the null hypothesis.

Let  $h_{im}(v(\tilde{\xi}_m), \tilde{\theta}) = \sum_{j=0}^J \frac{\partial \log P_{imj}(v(\tilde{\xi}_m), \tilde{\theta})}{\partial \tilde{\theta}} u_{imj}(\tilde{\theta})$  and let  $\overline{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta})] = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} E[h_{im}(v(\tilde{\xi}_m), \tilde{\theta})]$ . Again note that  $\overline{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta})] = 0$  when  $\tilde{\theta} = \tilde{\theta}_0$ , so the LM test uses  $h_{im}(v(\tilde{\xi}_m), \tilde{\theta})$  to detect a

violation of  $\tilde{\theta} = \tilde{\theta}_0$ , which is equivalent to the violation of the population moment condition (9). Further denote  $\tilde{\theta}_R^A = \text{plim}_{M \rightarrow \infty} \hat{\theta}_R$  under the alternative hypothesis against (7). Then, the main substantive condition for test consistency as we show in Theorem A.1 in Appendix A.3 is that we require

$$\lim_{M \rightarrow \infty} \|\bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]\| \neq 0 \quad (10)$$

for some function  $v(\tilde{\xi}_m)$  when prices are endogenous (i.e., the alternative hypothesis is true).<sup>15</sup> This means that for the test consistency  $v(\tilde{\xi}_m)$  should be chosen such that it can detect the nonzero correlation between  $\frac{\partial \log P_{imj}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)}{\partial \tilde{\theta}}$  and  $u_{imj}(\tilde{\theta}_R^A)$  under the alternative hypothesis. Theorem A.1 in Appendix A.3 spells out the exact conditions under which the LM test is consistent.

**Remark 1** *To further understand how the LM test works, without loss of generality, let the 0-th good be the outside good (i.e.,  $\xi_{m0} = 0$ ,  $x_{m0} = 0$ ,  $p_{m0} = 0$ , and  $U_{im0} = \epsilon_{im0}$ ). From  $u_{im0}(\tilde{\theta}) = -\sum_{j=1}^J u_{imj}(\tilde{\theta})$ , it follows that*

$$\begin{aligned} \sum_{j=0}^J \frac{\partial \log P_{imj}(\cdot, \tilde{\theta})}{\partial \tilde{\theta}} u_{imj}(\tilde{\theta}) &= \sum_{j=1}^J \frac{\partial \log P_{imj}(\cdot, \tilde{\theta})}{\partial \tilde{\theta}} u_{imj}(\tilde{\theta}) - \sum_{j=1}^J \frac{\partial \log P_{im0}(\cdot, \tilde{\theta})}{\partial \tilde{\theta}} u_{imj}(\tilde{\theta}) \\ &= \sum_{j=1}^J \frac{\partial \log(P_{imj}(\cdot, \tilde{\theta})/P_{im0}(\cdot, \tilde{\theta}))}{\partial \tilde{\theta}} u_{imj}(\tilde{\theta}). \end{aligned}$$

Define the part of the mean utility for the product  $j$  that is common across consumers as  $\delta_{mj}(\tilde{\xi}_m)$ . Then in particular we have

$$\begin{aligned} \sum_{j=0}^J \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)}{\partial (\lambda', \tilde{\gamma}', \tilde{\gamma}'_p)} u_{imj}(\hat{\theta}_R) &= \sum_{j=1}^J \frac{\partial \{P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)/P_{im0}(v(\hat{\xi}_m), \hat{\theta}_R)\}}{\partial \delta_{mj}} \\ &\quad \times \{v_j(\hat{\xi}_m)', x_{mj1}v_j(\hat{\xi}_m)', \dots, x_{mjK}v_j(\hat{\xi}_m)', p_{mj}v(\hat{\xi}_m)'\} u_{imj}(\hat{\theta}_R), \end{aligned} \quad (11)$$

which is the sum of the products of the omitted attributes (including interaction terms) and the residuals scaled by ‘‘hazard rates’’ (normalized w.r.t. the outside good) of purchases with respect to the mean utilities. Therefore the LM test rejects if the omitted attribute can significantly account for any of the purchase decision residual (i.e., if their correlation is significant). Indeed (11) simplifies to  $\sum_{j=0}^J \frac{\partial \log P_{imj}(v(\hat{\xi}_m), \hat{\theta}_R)}{\partial (\lambda', \tilde{\gamma}', \tilde{\gamma}'_p)} u_{imj}(\hat{\theta}_R) = \sum_{j=1}^J \{v_j(\hat{\xi}_m)', x_{mj1}v_j(\hat{\xi}_m)', \dots, x_{mjK}v_j(\hat{\xi}_m)', p_{mj}v(\hat{\xi}_m)'\} u_{imj}(\hat{\theta}_R)$  for the multinomial choice logit model (i.e.,  $\epsilon$  follows the type I extreme value distribution) without random coefficients. Therefore in the simple logit case the LM test seeks the correlation between the controls  $\{v_j(\hat{\xi}_m)', x_{mj1}v_j(\hat{\xi}_m)', \dots, x_{mjK}v_j(\hat{\xi}_m)', p_{mj}v(\hat{\xi}_m)'\}$  and  $u_{imj}(\hat{\theta}_R)$  to detect a violation of the moment condition (9).

## 6 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.

<sup>15</sup>Note that in (10) we allow for non-identically distributed data (e.g., the distributions of prices and/or the unobserved attributes are different across markets). If the data is iid across  $i$  and  $m$ , (10) simplifies to  $\|E[h_{im}(v(\tilde{\xi}_m), \theta_R^A)]\| \neq 0$ .

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

$$U_{im} = \beta_0 + \beta_i x_m + \beta_\xi \xi_m - \alpha p_m + \epsilon_{im},$$

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

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

where  $F(\beta_i)$  denotes the distribution function of the random coefficient.

We first examine four different monte carlo cases with only the specification for marginal cost varying across the cases (we add two cases of exogenous prices later). They 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:

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

In all specifications below  $x_m$  enters marginal cost, as do two other variables which do not affect demand:  $w_m$ , which is a cost-shifter observed by the researcher, and  $a_m$ , which is a cost shock that is not observed. Cases 1-4, in order, are given as:

$$\begin{aligned} [1] \quad mc_m &= 10 + x_m + \xi_m + w_m + a_m \\ [2] \quad mc_m &= 10 + x_m + w_m + a_m \\ [3] \quad mc_m &= 10 + \exp(x_m + \xi_m + w_m + a_m) \\ [4] \quad mc_m &= 10 + \exp(x_m + w_m + a_m). \end{aligned}$$

In two of the four specifications  $\xi_m$  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  $x_m, \xi_m, w_m$ , 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.

To evaluate whether the Wald test does not overreject the hypothesis of the exogenous price when the null hypothesis is true, we also generate the demand data without an unobserved attribute (i.e.  $\beta_\xi = 0$ ). We consider two additional specifications of the marginal costs:

$$\begin{aligned} [5] \quad mc_m &= 10 + x_m + w_m + a_m \\ [6] \quad mc_m &= 10 + \exp(x_m + w_m + a_m). \end{aligned}$$

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  $\xi_m$  and  $a_m$  in each market, instead seeing only  $x_m, w_m$ , and  $p_m$ , and OLS is used to estimate the residuals  $\hat{\xi}_m = p_m - \hat{E}[p_m|x_m, w_m]$ . While it makes the test less powerful we assume the researcher mistakenly adds a random coefficient, which is typical in many specifications. This adds a new parameter  $\sigma$ , the s.d. of the additional error term, and leads to estimation of the utility function given by

$$\tilde{U}_{im} = \beta_0 + \beta_i x_m + \lambda \hat{\xi}_m - \alpha p_m + \sigma \iota_m + \epsilon_{im}, \quad (12)$$

where  $\iota_m$  is distributed  $N(0, 1)$ .

The demand parameters are estimated using maximum likelihood in two ways. First we exclude  $\hat{\xi}_m$  from the specification to see how much bias the price endogeneity generates. We then follow the Wald approach from Section 3.1, including  $\hat{\xi}_m$  directly in the likelihood function and testing its coefficient  $\lambda$  for significance. For the standard errors from the Wald approach we bootstrap, repeating estimation 500 times (that is, on 500 different datasets, each with 200 markets) and using 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 1 summarizes the Monte Carlo data. Reported statistics are calculated for each sample of 200 markets, and then are averaged over the 500 iterations. The first four columns are the endogenous price specifications and the last two columns are the exogenous price specifications. Prices range from approximately 10 to 14. For the linear marginal cost specifications the 10%-90% range for shares is from 0.11 to 0.38 and for the exponential specifications they range from 0.02 to 0.19.  $Corr(p_m, \xi_m)$  is respectively 0.55, 0.15, 0.40, and 0.03 for endogenous prices and 0 for the exogenous price cases. The correlations are substantially higher in specifications 1 and 3 where the unobserved demand attribute affects costs.

Table 2 reports the parameter estimates for the approach without  $\hat{\xi}_m$  and for the Wald-type test specification from Section 3.1. The average of the parameter estimates and their standard deviation across the 500 iterations are reported first for the specification without  $\hat{\xi}_m$ . Next, the six parameters associated with the Wald test specification are reported, where the two new parameters are  $\sigma$ , the standard deviation of the additional error term, and  $\lambda$ , 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.

## 6.1 Endogenous Prices

### 6.1.1 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  $\hat{\xi}_m$  is included in the specification,  $\lambda$  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.

### **6.1.2 Case 2: Unobserved demand attribute does not affect (linear) marginal cost**

Case 2 has linear marginal costs, but  $\xi_m$  excluded, lowering the correlation between price and  $\xi_m$  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.  $\lambda$  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.

### **6.1.3 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.

### **6.1.4 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  $\xi_m$  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 has a price coefficient of 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.

## **6.2 Exogenous Prices**

### **6.2.1 Case 5: Linear marginal costs**

In this case both estimation approaches produce estimates very close to their true values. Thus, the Wald-type test does not reject the price exogeneity as expected. Here we are interested in the actual size of the test. Table 2 shows that the size of test is very close to the significance level. Under the null hypothesis (our DGP), the Wald-type test shows 1.2 % rejection of 500 cases with

0.01 significance level and 9.4 % rejection with 0.10 significance level. Therefore the Wald-type test does not overreject the null when the null hypothesis is true.

### 6.2.2 Case 6: Exponential marginal costs

In this case too both estimation approaches produce all estimates very close to their true values. Under the null hypothesis, the Wald-type test has 1.2 % rejection of 500 cases with 0.01 significance level and 9.0 % rejection with 0.10 significance level. Therefore the Wald-type test does not overreject in this case too.

## 7 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 3.1, where  $\hat{\xi}_m$  are included directly as new regressors in the likelihood function.

### 7.1 Data and Demand Specification

The data and specification are very similar to Goolsbee and Petrin (2004) and Petrin and Train (2010). 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

$$\tilde{U}_{imj} = \alpha p_{mj} + \sum_{g=2}^5 \theta_g p_{mj} 1_{ig} + \beta'_0 x_{mj} + \gamma'_j d_i + \sigma \nu_i c_j + \lambda_j \hat{\xi}_{mj} + \epsilon_{imj}.$$

$x_{mj}$  are the observed characteristics of the product and includes a product intercept term.  $\hat{\xi}_{mj}$  is the proxy and has  $\lambda_j$  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  $g$ .<sup>16</sup> Demographic variables for household  $i$  are given by  $d_i$  and enter each choice  $j$  with a separate coefficient vector  $\gamma_j$ . 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,  $\nu_i$  is an i.i.d. standard normal deviate, and  $\sigma$  is its

---

<sup>16</sup>The price coefficient for a household in the lowest income group is  $\alpha$  while that for a household in group  $g > 1$  is  $\alpha + \theta_g$ .

standard deviation, reflecting the degree of correlation among the non-antenna alternatives.  $\epsilon_{imj}$  is i.i.d. extreme value.<sup>17</sup>

In the Forrester survey, respondents reported the type of television medium that they have.<sup>18</sup> 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 each operator. Satellite prices do not vary geographically, and the price of antenna-only is assumed to be zero.<sup>19</sup> More complete details are provided in Goolsbee and Petrin (2004).

## 7.2 Estimation of the control

We estimate the controls 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 control term 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 control term is constructed in a similar manner.

---

<sup>17</sup>The error specification in Goolsbee and Petrin (2004) is more flexible; they use a multivariate normal specification in place of the logit error.

<sup>18</sup>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.

<sup>19</sup>For the price of satellite, we use \$50 per month plus an annual \$100 installation and equipment cost.

### 7.3 Results

Table 3 gives the estimated parameters and standard errors for the two approaches.<sup>20</sup> The first column of Table 3 gives the constrained model while the second column includes  $\hat{\xi}_{mj}$  as new regressors. The coefficients on the expanded basic and premium controls are reported first, followed by the base price coefficient. Both controls enter significantly (i.e. we reject exogenous prices) and with a positive sign, identifying the price endogeneity problem and suggesting that products with large proxies possess desirable attributes omitted from the specification. Inclusion of the controls also raises the magnitude of the estimated base price coefficient by 500%, from -0.02 to -0.09, consistent with bias associated with a positive correlation between the unobserved attributes and prices.

## 8 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 upwards, and evidence of it has been found in many applications spanning a wide range of markets and differing data types. The problem can be exacerbated when the interaction of price and the omitted attribute enters the consumer utility.

In this paper we develop a class of asymptotic tests for the price endogeneity. Our tests can allow for the non-additive separability between price and the omitted attribute. They are easy to implement in standard regression packages and have reasonable power against the "exogenous prices" hypothesis. The tests include control function type tests of specification.

A major advantage of our testing approaches is their simplicity. They all use least squares or nonparametric regression in an initial stage to obtain the proxies, followed by a second stage of likelihood function maximization. The control function test includes the controls directly in the maximization of the likelihood function and tests for their significance.

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

---

<sup>20</sup>We bootstrap to account for the additional variance from estimating the expected price. Specifically, 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 demand and supply factors from the 172 markets. We then estimate  $\hat{E}[p_{mj}|Z_m]$  and calculate the implied (new) controls, and then re-estimate the model with these controls. 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.

(Chintagunta, Dube, and Goh (2005)). Our Monte Carlo 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 almost every case.

**Table 1**  
Summary of Data Generated in Monte Carlo Samples

Case	[1]	[2]	[3]	[4]	[5]	[6]
<b>True values:</b> $\alpha = 1, \beta_0 = 10, \beta_x = 1, \sigma_x = 0.5$						
Price Range						
Mean	11.31	11.33	12.75	12.57	11.32	12.57
10%-90%	10.17-12.45	10.36-12.28	11.41-14.60	11.49-14.03	10.37-12.27	11.49-14.03
Share Range						
Mean	0.22	0.23	0.09	0.10	0.23	0.10
10%-90%	0.11-0.34	0.10-0.38	0.01-0.16	0.02-0.19	0.12-0.35	0.02-0.17
% of markets with						
shares < 0.01	0.0%	0.0%	7.4%	5.7%	0.0%	5.4%
shares < 0.10	5.8%	10.0%	57.1%	53.5%	5.8%	51.0%
<b>Corr</b> ( $p_m, \xi_m$ )	0.55	0.15	0.40	0.03	0.00	0.00
Marginal Costs						
Linear	Yes	Yes	No	No	Yes	No
Exponential	No	No	Yes	Yes	No	Yes
Include $\xi_m$	Yes	No	Yes	No	No	No

Cases [1]-[4] are with an omitted attribute ( $\beta_\xi = 1$ ) and Cases [5]-[6] are with no omitted attribute ( $\beta_\xi = 0$ ). Reported numbers are the statistics for each of the samples with 200 markets averaged over the 500 Monte Carlo iterations.

**Table 2**  
 Constrained Model and Wald-type Test Specification for Monte Carlo Data  
 200 markets, 20 observations per market, 500 iterations

<b>True values:</b> $\alpha = 1, \beta_0 = 10, \beta_x = 1, \sigma_x = 0.5$						
Case	[1]	[2]	[3]	[4]	[5]	[6]
Constrained Model						
$\alpha$	0.51 (0.03)	0.74 (0.06)	0.63 (0.03)	0.86 (0.05)	1.00 (0.01)	1.00 (0.01)
$\beta_0$	4.56 (0.38)	7.17 (0.69)	5.52 (0.43)	8.40 (0.66)	9.97 (0.09)	9.98 (0.09)
$\beta_x$	0.51 (0.06)	0.74 (0.09)	0.65 (0.07)	0.88 (0.09)	1.00 (0.02)	1.00 (0.02)
$\sigma_x$	0.23 (0.45)	0.24 (0.49)	0.12 (0.33)	0.22 (0.41)	0.34 (0.25)	0.44 (0.13)
Wald Test/Control Function						
$\alpha$	0.99 (0.09)	1.01 (0.26)	0.81 (0.08)	0.94 (0.09)	1.00 (0.01)	1.00 (0.01)
$\beta_0$	9.90 (1.11)	10.13 (2.48)	7.69 (0.95)	9.20 (1.02)	9.98 (0.12)	9.98 (0.09)
$\beta_x$	0.99 (0.12)	1.01 (0.25)	0.99 (0.13)	1.00 (0.14)	1.00 (0.02)	1.00 (0.02)
$\sigma_x$	0.44 (0.23)	0.52 (0.40)	0.26 (0.19)	0.41 (0.25)	0.41 (0.11)	0.45 (0.09)
$\sigma$	0.28 (0.20)	0.53 (0.69)	0.55 (0.35)	0.60 (0.45)	0.05 (0.04)	0.05 (0.04)
<b>Test statistic: <math>\lambda</math></b>	0.62 (0.10)	0.39 (0.15)	0.25 (0.07)	0.08 (0.07)	0.00 (0.01)	0.00 (0.01)
Rejection rate at:						
size=0.01	100%	99.2%	99.6%	85.2%	1.2%	1.2%
size=0.10	100%	99.4%	99.6%	93.2%	9.4%	9.0%

Cases [1]-[4] are with an omitted attribute ( $\beta_\xi = 1$ ) and Cases [5]-[6] are with no omitted attribute ( $\beta_\xi = 0$ ). 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  $\lambda$  under the null hypothesis (estimates not reported here).

**Table 3**  
TV Reception Choice  
Constrained Approach and the Wald-type Test Specification

Explanatory variable	Constrained Approach (Standard errors in parentheses)	Wald-type Test Specification
<b>Control for expanded-basic cable price (2)</b>		.0805 (.0416)
<b>Control for premium cable price (3)</b>		.0873 (.0418)
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)
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.

## Appendix

### A Asymptotic Distributions of Test Statistics

#### A.1 Proof of Lemma 5.1

**Proof.** The unconstrained ML estimator solves the first order condition  $\frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_U)}{\partial \theta} = 0$ . Define  $\hat{\Gamma}(\tilde{\theta}, \pi) = -\frac{\partial \log L_N(\tilde{\xi}_m(\pi), \tilde{\theta})}{\partial \tilde{\theta} \partial \theta'}$ ,  $\Gamma_0(\tilde{\theta}, \pi) = \lim_{M \rightarrow \infty} E[-\frac{\partial \log L_N(\tilde{\xi}_m(\pi), \tilde{\theta})}{\partial \tilde{\theta} \partial \theta'}$ ],  $\hat{\Gamma}(\tilde{\theta}) = \hat{\Gamma}(\tilde{\theta}, \hat{\pi})$ , and  $\Gamma_0 = \Gamma_0(\tilde{\theta}_0, \pi_0)$ . The asymptotic distribution of  $\sqrt{M}(\hat{\theta}_U - \tilde{\theta}_0)$  is obtained from the asymptotic expansion of

$$\sqrt{M}(\hat{\theta}_U - \tilde{\theta}_0) = \hat{\Gamma}^{-1}(\tilde{\theta}^*) \left\{ \sqrt{M} \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} + \sqrt{M} \left( \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} - \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} \right) \right\} + o_p(1), \quad (13)$$

which is obtained from the element-by-element mean value expansions of the first order condition  $\frac{\partial \log L_N(\hat{\xi}_m, \hat{\theta}_U)}{\partial \theta} = 0$  around  $\tilde{\theta}_0$  where  $\tilde{\theta}^*$  lies between  $\hat{\theta}_U$  and  $\tilde{\theta}_0$ . Therefore the asymptotic variance of  $\sqrt{M}(\hat{\theta}_U - \tilde{\theta}_0)$  contains two variance terms that are from (i) the ML estimation and (ii) the estimation of the controls.

Note that under Assumptions 5.1, 5.2 (i)(a), (v), and (vi), by the uniform law of large numbers and the continuity, we obtain

$$\hat{\Gamma}(\tilde{\theta}^*) \rightarrow_p \Gamma_0 \quad (14)$$

because

$$\begin{aligned} \left\| \hat{\Gamma}(\tilde{\theta}^*) - \Gamma_0 \right\| &\leq \left\| \hat{\Gamma}(\tilde{\theta}^*) - \Gamma_0(\tilde{\theta}^*, \hat{\pi}) \right\| + \left\| \Gamma_0(\tilde{\theta}^*, \hat{\pi}) - \Gamma_0 \right\| \\ &\leq \sup_{\tilde{\theta} \in \tilde{\Theta}_0, \pi \in \Pi_0} \left\| \hat{\Gamma}(\tilde{\theta}, \pi) - \Gamma_0(\tilde{\theta}, \pi) \right\| + \left\| \Gamma_0(\tilde{\theta}^*, \hat{\pi}) - \Gamma_0 \right\| = o_p(1) \end{aligned} \quad (15)$$

where the first term in (15) converges to zero by the uniform LLN under Assumption 5.2 (vi) and the second term in (15) converges to zero because of the continuity of  $\Gamma_0$  (which is implied by Assumption 5.2 (v) and (vi) due to the dominated convergence theorem) and because  $(\tilde{\theta}^*, \hat{\pi}) \rightarrow_p (\tilde{\theta}_0, \pi_0)$ .

To derive the variance term due to the second term inside  $\{\cdot\}$  bracket in (13), we approximate the second term in (13) using a first order mean value expansion,

$$\begin{aligned} &\sqrt{M} \left( \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} - \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} \right) \quad (16) \\ &= \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j''=0}^J Y_{imj''} \sum_{j'=1}^J \frac{\partial^2 \log P_{imj''}(v(\tilde{\xi}_m(\hat{\pi}^*)), \tilde{\theta}_0)}{\partial \tilde{\theta} \partial v_{j'}(\tilde{\xi}_m)} \sum_{j=1}^J \frac{\partial v_{j'}(\tilde{\xi}_m(\hat{\pi}^*))}{\partial \tilde{\xi}_{mj}} \frac{\partial \tilde{\xi}_{mj}(\hat{\pi}^*)}{\partial \pi_j} \sqrt{M}(\hat{\pi}_j - \pi_{j0}) \\ &\quad + o_p(1) \end{aligned}$$

where  $\hat{\pi}^*$  lies between  $\hat{\pi}$  and  $\pi_0$ .

Define  $\Lambda_j^{N,M}(\tilde{\theta}, \tilde{\xi}_m) = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} \sum_{j''=0}^J Y_{imj''} \sum_{j'=1}^J \frac{\partial^2 \log P_{imj''}(v(\tilde{\xi}_m), \tilde{\theta})}{\partial \tilde{\theta} \partial v_{j'}(\tilde{\xi}_m)} \frac{\partial v_{j'}(\tilde{\xi}_m)}{\partial \tilde{\xi}_{m_j}} \frac{\partial \tilde{\xi}_{m_j}}{\partial \pi_j}$ . Then we can rewrite (16) as

$$\begin{aligned} & \sqrt{M} \left( \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} - \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} \right) \\ &= \sum_{j=1}^J \Lambda_j^{N,M}(\tilde{\theta}_0, \tilde{\xi}_m(\hat{\pi}^*)) \varpi_j + o_p(1) \rightarrow_d \mathcal{N}(0, \sum_{j,k} \Lambda_{0j} \text{Cov}(\varpi_j, \varpi_k) \Lambda'_{0k}) = \mathcal{N}(0, V_1) \end{aligned}$$

where  $\Lambda_{0j} = \lim_{M \rightarrow \infty} E[\Lambda_j^{N,M}(\tilde{\theta}_0, \tilde{\xi}_m)]$  and the asymptotic normality result holds by the continuous mapping theorem and by the Lindeberg-Feller CLT under Assumption 5.1. In the above we can show  $\Lambda_j^{N,M}(\tilde{\theta}_0, \tilde{\xi}_m(\hat{\pi}^*)) \rightarrow_p \Lambda_{0j}$  by following similar steps to (14) under Assumptions 5.1 and 5.2.

Then by  $\sqrt{M} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} \rightarrow_d \mathcal{N}(0, \Gamma_0 \cdot \lim_{M \rightarrow \infty} \frac{M}{N})$  (due to the Lindeberg-Feller CLT under Assumption 5.2 (vii)), (14), the Slutsky theorem, and the continuous mapping theorem, from (13) we obtain

$$\sqrt{M}(\hat{\theta}_U - \tilde{\theta}_0) \rightarrow_d \mathcal{N}(0, \Gamma_0^{-1}(\Gamma_0 \cdot \lim_{M \rightarrow \infty} \frac{M}{N} + V_1)\Gamma_0^{-1}). \quad (17)$$

■

## A.2 Asymptotic Distribution of the LM test

**Proof.** We show that the feasible LM test statistic has the same asymptotic distribution with the corresponding Wald test statistic.

Element-by-element mean value expansions of  $\frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_R)}{\partial \tilde{\theta}}$  around  $\tilde{\theta}_0$  yield

$$\sqrt{M} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_R)}{\partial \tilde{\theta}} = \sqrt{M} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} + \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}^*)}{\partial \tilde{\theta} \partial \tilde{\theta}'} \sqrt{M}(\tilde{\theta}_R - \tilde{\theta}_0) + o_p(1) \quad (18)$$

where  $\tilde{\theta}^*$  lies between  $\tilde{\theta}_R$  and  $\tilde{\theta}_0$ . Note that  $H(\tilde{\theta}_R - \tilde{\theta}_0) = 0$  by construction of  $H$ . Write  $\hat{\Gamma}(\tilde{\theta}^*) = -\frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}^*)}{\partial \tilde{\theta} \partial \tilde{\theta}'}$ . Then by multiplying  $H\hat{\Gamma}(\tilde{\theta}^*)^{-1}$  to both sides of (18), it follows that

$$\begin{aligned} & \sqrt{M} H \hat{\Gamma}(\tilde{\theta}^*)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_R)}{\partial \tilde{\theta}} \\ &= \sqrt{M} H \hat{\Gamma}(\tilde{\theta}^*)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} + o_p(1) = \sqrt{M} H \Gamma_0^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} + o_p(1) \\ &= \sqrt{M} H \Gamma_0^{-1} \left( \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} + \left( \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} - \frac{\partial \log L_N(\tilde{\xi}_m, \tilde{\theta}_0)}{\partial \tilde{\theta}} \right) \right) + o_p(1) \end{aligned}$$

by the similar argument with (14) and the Slutsky theorem. Therefore, under the null hypotheses (7),  $\sqrt{M} H \hat{\Gamma}(\tilde{\theta}^*)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_R)}{\partial \tilde{\theta}}$  follows the same asymptotic distribution with  $\sqrt{M} H(\hat{\theta}_U - \tilde{\theta}_0)$  and we obtain from (17)

$$\sqrt{M} H \hat{\Gamma}(\tilde{\theta}_R)^{-1} \frac{\partial \log L_N(\hat{\xi}_m, \tilde{\theta}_R)}{\partial \tilde{\theta}} \rightarrow_d \mathcal{N}(0, H \Gamma_0^{-1}(\Gamma_0 \lim_{M \rightarrow \infty} \frac{M}{N} + V_1)\Gamma_0^{-1} H') \quad (19)$$

by the Lindeberg-Feller CLT and the continuous mapping theorem and because  $\widehat{\Gamma}(\widehat{\theta}_R) \rightarrow_p \Gamma_0$ ,  $\widehat{\Gamma}(\widehat{\theta}^*) \rightarrow_p \Gamma_0$ . Then by (19),  $\widehat{\Gamma}(\widehat{\theta}_R) \rightarrow_p \Gamma_0$ ,  $\widehat{V}_1(\widehat{\theta}_R) \rightarrow_p V_1$ , and the continuous mapping theorem it follows that

$$\begin{aligned} \widetilde{LM}_{M,N} &= N \frac{\partial \log L_N(\widehat{\xi}_m, \widehat{\theta}_R)}{\partial \theta'} \widehat{\Gamma}(\widehat{\theta}_R)^{-1} H' \widetilde{V}_{M,N}^{-1}(\widehat{\theta}_R) H \widehat{\Gamma}(\widehat{\theta}_R)^{-1} \frac{\partial \log L_N(\widehat{\xi}_m, \widehat{\theta}_R)}{\partial \theta} \\ &= \left\{ \sqrt{M} H \widehat{\Gamma}(\widehat{\theta}_R)^{-1} \frac{\partial \log L_N(\widehat{\xi}_m, \widehat{\theta}_R)}{\partial \theta} \right\}' \cdot \left\{ H \widehat{\Gamma}(\widehat{\theta}_R)^{-1} (\widehat{\Gamma}(\widehat{\theta}_R) \frac{M}{N} + \widehat{V}_1(\widehat{\theta}_R)) \widehat{\Gamma}(\widehat{\theta}_R)^{-1} H' \right\}^{-1} \\ &\times \left\{ \sqrt{M} H \widehat{\Gamma}(\widehat{\theta}_R)^{-1} \frac{\partial \log L_N(\widehat{\xi}_m, \widehat{\theta}_R)}{\partial \theta} \right\} \rightarrow_d \chi^2(\dim((\lambda', \gamma', \gamma'_p)')). \end{aligned}$$

Therefore,  $\widetilde{LM}_{M,N}$  follows the same asymptotic distribution of the Wald test statistic  $\widetilde{T}_{M,N}$  under the null hypothesis of the price exogeneity. ■

### A.3 Consistency of the LM test

Define  $\bar{h}_{M,N}(v(\tilde{\xi}_m), \tilde{\theta}) = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} h_{im}(v(\tilde{\xi}_m), \tilde{\theta})$  and  $\bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta})] = \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} E[h_{im}(v(\tilde{\xi}_m), \tilde{\theta})]$ .

Assume the following to show the consistency of the LM test.

**Assumption A.1** (i)  $\widehat{\theta}_R \rightarrow_p \tilde{\theta}_R^A$  under the alternative hypothesis against (7); (ii)-(viii) each corresponding Assumption 5.2 (ii) to (viii) holds replacing  $\tilde{\theta}_0$  with  $\tilde{\theta}_R^A$  and  $\tilde{\Theta}_0$  with  $\tilde{\Theta}_A$  where  $\tilde{\Theta}_A$  denotes a neighborhood of  $\tilde{\theta}_R^A$ .

**Theorem A.1** Suppose Assumptions 5.1 and A.1 hold. Then the LM test  $\widetilde{LM}_{M,N}$  is consistent under  $\lim_{M \rightarrow \infty} |\bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]| \neq 0$ .

**Proof.** Under the alternative hypothesis against (7), using a mean value expansion, we obtain

$$\sqrt{M} \bar{h}_{M,N}(v(\widehat{\xi}_m), \widehat{\theta}_R) = \frac{1}{\sqrt{N}} \sum_{m=1}^M \sum_{i=1}^{N_m} (h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A) - E[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]) \times \frac{\sqrt{M}}{\sqrt{N}} \quad (20)$$

$$+ \sqrt{M} (\bar{h}_{M,N}(v(\widehat{\xi}_m), \widehat{\theta}_R) - \bar{h}_{M,N}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)) \quad (21)$$

$$+ \frac{\partial \bar{h}_{M,N}(v(\widehat{\xi}_m), \tilde{\theta}^{A*})}{\partial \theta'} \sqrt{M} (\widehat{\theta}_R - \tilde{\theta}_R^A) + \sqrt{M} \bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)], \quad (22)$$

where  $\tilde{\theta}^{A*}$  lies between  $\widehat{\theta}_R$  and  $\tilde{\theta}_R^A$ . Now we analyze each term one by one below.

For (21) applying the mean value expansion around  $\pi_0$ , we obtain

$$\sqrt{M} (\bar{h}_{M,N}(v(\widehat{\xi}_m), \widehat{\theta}_R) - \bar{h}_{M,N}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)) = \frac{\partial \bar{h}_{M,N}(v(\widehat{\xi}_m(\pi^*), \tilde{\theta}_R^A)}{\partial \pi'} \sqrt{M} (\widehat{\pi} - \pi_0) \quad (23)$$

where  $\pi^*$  lies between  $\widehat{\pi}$  and  $\pi_0$ . Let  $\Gamma_{\pi A} = \lim_{M \rightarrow \infty} E[\frac{\partial \bar{h}_{M,N}(v(\widehat{\xi}_m(\pi^*), \tilde{\theta}_R^A)}{\partial \pi}]$ . We then have

$$\sqrt{M}(\bar{h}_{M,N}(v(\hat{\xi}_m), \tilde{\theta}_R^A) - \bar{h}_{M,N}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)) = \Gamma'_{\pi_A} \varpi + o_p(1)$$

by Assumption 5.1 and because under  $E[\sup_{\pi \in \Pi_0} \|\frac{\partial h_{im}(v(\tilde{\xi}_m(\pi)), \tilde{\theta}_R^A)}{\partial \pi}\|^2] < \infty$  for  $m$  (which holds under Assumption A.1 (viii)), we have  $\frac{\partial \bar{h}_{M,N}(v(\tilde{\xi}_m(\pi^*)), \tilde{\theta}_R^A)}{\partial \pi} \rightarrow_p \Gamma_{\pi_A}$  by the uniform Law of Large numbers and because  $E[\frac{\partial \bar{h}_{M,N}(v(\tilde{\xi}_m(\pi)), \tilde{\theta}_R^A)}{\partial \pi}]$  is continuous at  $\pi = \pi_0$  (which is implied by Assumption A.1 (viii) due to the dominated convergence theorem).

Next to analyze the first term in (22) let the inverse of the asymptotic variance matrix of the unconstrained estimator  $\hat{\theta}_U$  be  $B = \Gamma_0(\Gamma_0 \cdot \lim_{M \rightarrow \infty} \frac{M}{N} + V_1)^{-1} \Gamma_0$  and define a matrix  $\mathcal{M} = I - B^{-1/2} H'(HB^{-1}H')^{-1}HB^{-1/2}$ . Then the asymptotic distribution of the constrained estimator  $\hat{\theta}_R$  is given by  $\sqrt{M}(\hat{\theta}_R - \tilde{\theta}_R^A) \rightarrow_d \mathcal{N}(0, B^{-1/2} \mathcal{M} B^{-1/2}) \equiv \mathcal{Z}_{2A}$  (see p. 2217-2220 of Newey and McFadden (1994)). Then we obtain

$$\frac{\partial \bar{h}_{M,N}(v(\hat{\xi}_m), \tilde{\theta}^{A*})}{\partial \theta'} \sqrt{M}(\hat{\theta}_R - \tilde{\theta}_R^A) = \Gamma'_{\theta_A} \mathcal{Z}_{2A} + o_p(1)$$

where  $\sqrt{M}(\hat{\theta}_R - \tilde{\theta}_R^A) \rightarrow_d \mathcal{Z}_{2A}$  and  $\Gamma_{\theta_A} = \lim_{M \rightarrow \infty} E[\frac{\partial \bar{h}_{M,N}(v(\hat{\xi}_m), \tilde{\theta}^{A*})}{\partial \theta}]$  because under Assumption A.1 (vi)-(v), we have  $\frac{\partial \bar{h}_{M,N}(v(\hat{\xi}_m), \tilde{\theta}^{A*})}{\partial \theta} \rightarrow_p \Gamma_{\theta_A}$  by the uniform Law of Large numbers, because  $E[\frac{\partial \bar{h}_{M,N}(v(\tilde{\xi}_m(\pi)), \tilde{\theta}_R^A)}{\partial \theta}]$  is continuous at  $\tilde{\theta}_R = \tilde{\theta}_R^A$  and  $\pi = \pi_0$  (which is implied by Assumption A.1 (vi)-(v) due to the dominated convergence theorem), and because  $\tilde{\theta}^{A*} \rightarrow_p \tilde{\theta}_R^A$  and  $\hat{\pi} \rightarrow_p \pi_0$ .

Next we consider the term in (20). Note that under  $E[\|h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)\|^4] < \infty$  for all  $m$  (which holds under Assumption A.1 (vii)), by the Lindeberg-Feller CLT, we have

$$\sum_{m=1}^M \sum_{i=1}^{N_m} (h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A) - E[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]) / \sqrt{N} \rightarrow_d \mathcal{Z}_{1A} \equiv \mathcal{N}(0, V_{hA})$$

where  $V_{hA} = \lim_{M \rightarrow \infty} \frac{1}{N} \sum_{m=1}^M \sum_{i=1}^{N_m} E[\{h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A) - E[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]\} \cdot \{h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A) - E[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]\}']$ . Combining these results, we obtain

$$\begin{aligned} \sqrt{M} \bar{h}_{M,N}(v(\hat{\xi}_m), \hat{\theta}_R) &= \mathcal{Z}_{1A} \frac{\sqrt{M}}{\sqrt{N}} + \Gamma'_{\pi_A} \varpi + \Gamma'_{\theta_A} \mathcal{Z}_{2A} + \sqrt{M} \bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)] + o_p(1) \\ &= O_p(1) + \sqrt{M} \bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]. \end{aligned}$$

Therefore, we obtain  $\|\sqrt{M} \bar{h}_{M,N}(v(\hat{\xi}_m), \hat{\theta}_R)\| \rightarrow \infty$  if  $\lim_{M \rightarrow \infty} \|\bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]\| \neq 0$  under the alternative against (7). This implies  $\bar{L} \bar{M}_{M,N} \rightarrow \infty$  under the alternative if  $\lim_{M \rightarrow \infty} \|\bar{E}[h_{im}(v(\tilde{\xi}_m), \tilde{\theta}_R^A)]\| \neq 0$ . Therefore the LM test for price endogeneity is consistent. ■

## References

- Altonji, J., and R. Matzkin (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053-1102.
- Andrews, D. (1997), "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097-1128.
- Berry, S. (1994), "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242-262.
- Berry, S., A. Gandhi, and P. Haile (2011), "Connected Substitutes and Invertibility of Demand," Working Paper.
- Berry, S., and P. Haile (2010), "Identification in Differentiated Products Markets Using Market Level Data," Working Paper.
- Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-889.
- Bierens, H. (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.
- Chen, X., and Y. Fan (1999), "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373-401.
- Chintagunta, P., J. Dube, and K. Goh (2005), "Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models," *Management Science*, 51(5), 832-849.
- Crawford, G. (2000), "The Impact of the 1992 Cable Act on Household Demand and Welfare," *RAND Journal of Economics*, 31, 422-449.
- Dubé, J.P., J. Fox, and C.L. Su (2012), "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80(5), 2231-2267.
- Fan, Y., and Q. Li (1996), "Consistent Model Specification Tests: Omitted Variables, Parametric and Semiparametric Functional Forms," *Econometrica*, 64, 865-890.
- Gandhi, A., K. Kim, and A. Petrin (2011), "Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error," Working Paper.
- Goolsbee, A., and A. Petrin (2004), "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," *Econometrica*, 72(2), 351-382.

- Hardle, W., and E. Mammen (1993), "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926-1947.
- Hausman, J. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1272.
- Hausman, J. (1997), "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by R. Gordon, and T. Bresnahan. University of Chicago Press, Chicago.
- Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931-959.
- Heckman, J., and R. Robb (1985), "Alternative Methods for Evaluating the Impacts of Interventions: An Overview," *Journal of Econometrics*, 30, 239-267.
- Horowitz, J., and V. Spokoiny (2001), "An Adaptive, Rate Optimal Test of a Parametric Mean Regression Model against a Nonparametric Alternative," *Econometrica*, 69, 599-631.
- 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.
- Kim, K., and A. Petrin (2010), "Control Function Corrections for Unobserved Factors in Differentiated Product Models," Working Paper.
- Kitamura, Y., G. Tripathi, and H. Ahn (2004), "Empirical Likelihood-based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667-1714.
- Knittel, C. R. and K. Metaxoglou (2008), "Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings," Working Paper, MIT.
- Li, Q., C. Hsiao, and J. Zinn (2003), "Consistent Specification Tests for Semiparametric/ Nonparametric Models Based on Series Estimation Methods," *Journal of Econometrics*, 112, 295-325.
- Matzkin, R. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71.
- Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307-342.
- Newey, W. (1997), "Convergence Rates and Asymptotic Normality of Series Estimators," *Journal of Econometrics*, 79, 147-168.

- Newey, W., and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," *The Handbook of Econometrics*, Volume 4, Elsevier.
- 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 (2010), "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47, 370-379.
- 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.
- Song, K. (2010), "Testing Semiparametric Conditional Moment Restrictions Using Conditional Martingale Transforms," *Journal of Econometrics*, 154, 74-84.
- Stinchcombe, M., and H. White (1998), "Consistent Specification Testing When the Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295-325.
- Trajtenberg, M. (1989), "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 97, 444-479.
- Villas-Boas, J., and R. Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45, 1324-1338.