Estimating Price Elasticities in Differentiated Product Demand Models with Endogenous Characteristics

Daniel A. Ackerberg[*] Gregory S. Crawford
UCLA University of Arizona

March 27, 2009
Preliminary and Incomplete - Please do not distribute

Abstract

Empirical models of differentiated product demand have typically allowed price to be endogenous, but proceed under the assumption that observed product characteristics are exogenous, i.e. uncorrelated with unobserved components of demand. This paper shows that such an assumption may not be necessary to obtain consistent estimates of price elasticities. We show that whether this is the case depends on properties of the instrument or instruments used for price. Since these properties are testable, this result has interesting implications on an applied researcher's choice of price instruments. In the case where one cannot find an instrument that satisfies these properties, one can often bound the potential bias in estimated price elasticities due to endogenous product characteristics. Our ideas also lead to interesting thoughts about what sorts of variables would ideally like to have as price instruments. Lastly, we apply these ideas to data on demand for cable television, obtaining estimates of price elasticities that are in fact robust to endogenous product characteristics.

[*]Thanks to Jin Hahn and Joris Pinkse for comments and helpful discussion. All errors are our own.

1 Introduction

The recent literature in empirical IO has devoted significant effort to developing methodology for estimating demand systems in differentiated product markets. This is understandable since 1) many, if not most, markets are characterized by product differentiation, and 2) demand systems are a crucial component of many interesting IO questions. For example, an analysis of the price effects of mergers will depend on estimates of own and cross-price elasticities, which will typically come from an estimated demand system. An estimated demand system is also crucial for, e.g., measuring elasticities w.r.t. product characteristics, computing the welfare effects of new products or price changes, and creating optimal price indices. They can also be important indirect inputs in answering many other interesting IO questions. The general idea here is that if one is going to address general IO issues, one typically needs to have the demand system right.

Starting with Bresnahan (1987) and Berry, Levinsohn, and Pakes (1995, henceforth BLP), an important contribution of this methodology has been to base these demand systems on characteristics of the various products. This reduces the dimensionality of the system from the number of products available on the market (which can be very large, e.g. automobiles or computers) to the dimension of product characteristic space, making estimation feasible. Another key issue that has been addressed is that of price endogeneity. Given prices are choice variables of firms, it is likely that they will respond to components of demand that are unobserved to the econometrician, creating an endogeneity problem. Estimation has typically proceeded by trying to find appropriate "instruments" for price in order to solve this endogeneity problem.

In contrast, both the developers and appliers of this methodology have admittedly ignored potential endogeneity of product characteristics. In other words, practitioners have relied on the assumption that product characteristics are exogenous. Just like price, product characteristics are typically choice variables of firms, and as such one might worry that they are actually correlated with unobserved components of demand.

So why have existing studies relied on the assumption that product characteristics are exogenous? One reason is that it is likely in many intances that product characteristics are less endogenous than price. The idea here is that since price is often a more flexible decision than are product characteristics, it might be expected to be more correlated with unobserved demand components. On the other hand, it would certainly be preferable to allow for endogenous product characteristics as well. A second reason is that currently there are no particularly attractive alternatives. One possibility is to instrument for product characteristics as well as price, but this would require finding instruments for each product characteristics. It is already hard enough finding credible instruments for price. In addition, when one allows for the possibility of unobserved product characteristics, it is hard to imagine being able to find such instrumental variables. They would need to be correlated with observed product characteristics, but uncorrelated with unobserved product characteristics. It is unclear what variables would satisfy such a requirement.

Another possibility, suggested by BLP (and applied recently in Sweeting (2008) and Fan (2009)), involves setting up moments in terms of *innovations* in demand unobservables rather than the demand unobservables themselves. This, when accompanied by an assumption on what firms' information sets contain at the time when product characteristics are decided, can provide consistent estimates even when product characteristics are endogenous. BLP admit that this solution, which in practice involves psuedo-differencing the data, can be demanding on the data and we have yet to see it applied.

Lastly, one could imagine actually modelling firms' choices of product characteristics. Crawford and Shum (2006) take this approach. While this might be the most appealing approach to the problem conceptually, it ends up being quite complicated in practice. As such, Crawford and Shum are forced to restrict attention to monopoly markets (although multiproduct) where there is only a one dimensional product characteristic. Their approach would be much harder if not impossible to apply to either oligopoly markets or markets with multidimensional product characteristics. Pioner (2008) studies non-parametric identification in a similar model.

Our paper takes a different, simpler, approach to the problem of endogenous product characteristics. We start with the observation that one does not always need to estimate causal effects of changing product characteristics - for many questions, e.g. most short-run antitrust questions, one is primarily interested in own and cross price elasticities. Given this, we ask the following question: under what conditions will we get consistent estimates of price elasticities, *even* if product characteristics are endogenous?

Some simple econometric derivations show that under particular assumptions on our price instruments, standard estimation procedures (e.g. BLP) *do in fact* return consistent estimates of price elasticities, *even* if product characteristics are endogenous. These assumptions involve correlations between the instruments and the endogenous product characteristics, and perhaps most important, are testable. As a result, one can 1) test if our estimates are consistent even if product characteristics are endogenous, and 2) if not, perhaps appropriately choose price instruments to generate estimates that *are* robust to endogenous product characteristics. The analysis also sheds light on what types of data generating processes would generate instruments that satisfy this condition. In some cases where we cannot obtain consistent estimates of price elasticities, we can bound the bias in these estimates.

We end by illustrating our methodology using a dataset on demand for cable television. The data includes information from almost 5000 cable systems across the US. Our two key explanatory variables are price and the product characteristic "number of cable networks" offered. We have a number of potential price instruments, and we are able to test the above conditions for each of the various instruments. One of these instruments satisties this condition, and thus by using this instrument we are able to consistently estimate price elasticities allowing for endogenous product characteristics. Interestingly, this instrument also gives more reasonable price elasticities than the others.

2 Econometric Preliminaries

We start by discussing some relatively simple econometric results that are relevant for our treatment of endogenous product characteristics. These results all concern instrumental variables estimation of casual effects in the presence of covariates. A key difference from the standard treatment of instrumental variables is that we will consider a situation where one is *not* interested in estimating the casual effects of the covariates (on the dependent variable). This allows us to use identification conditions that are slightly different than the standard IV conditions. We later argue that these alternative identification conditions are particularly useful in a situation where product characteristics are endogenous. Another interesting attribute of these alternative identification conditions is that they are partially testable. We start by showing these ideas in a linear situation, and then generalize the ideas to the non-parametric model of Chernozhukov, Imbens, and Newey (2006).

2.1 Linear Model

Consider a linear model of the form

$$y_i = \beta_1 x_i + \beta_2 p_i + \epsilon_i \tag{1}$$

β_1 and β_2 respectively measure the causal effects of observables x_i and p_i on y_i. ϵ_i represents unobservables that also affect y_i. Looking ahead to our application, one might interpret (1) as the demand curve for a product whose characteristics and price vary across markets - p_i is the price in market i, x_i is an L-vector of the product's characteristics in market i, and y_i is quantity demanded. ϵ_i are unobservables that could represent either characteristics of the product that are not observed by the econometrician or demand shocks in market i.

Throughout, we will assume that p_i is potentially correlated with ϵ_i, i.e. that it is endogenous. We will also consider the possibility that x_i is endogenous. As mentioned in the introduction and above, a key distinction between x_i and p_i is that we assume that we are primarily interested in estimating the causal effect of p_i on y_i. In contrast we are less interested or not interested in the causal effects of x_i on y_i. We assume that we observe z_i, a potential instrument for p_i. In the demand context, one can think of z_i as a cost shifter. In contrast, we assume we do not have outside instruments for the covariates x_i. WLOG, all variables are assumed mean-zero.

Consider IV estimation of (1) using (x_i, z_i) as instruments for (x_i, p_i). Aside from regularity and rank conditions, the typical assumptions made to ensure identification of the causal effect β_2 are

Assumption L1: $E[\epsilon_i z_i] = 0, E[\epsilon_i x_i] = 0$

Note that for simplicity we are considering the necessity for the instrument z_i to be correlated with p_i (conditional on x_i) as a "rank" condition. This will be implicitly assumed throughout.

There are two compoments of Assumption $(L1)$. The first states that z_i is a valid instrument for p_i, i.e. it is uncorrelated with the residual. As is well known, without outside instruments, $E[\epsilon_i x_i] = 0$ is also generally necessary for identification of β_2. Even if $E[\epsilon_i z_i] = 0$, any correlation between ϵ_i and x_i will generally render IV estimates of β_2 inconsistent.This "transmitted bias" in analagous to that when one uses OLS when one regressor is exogenous and another is endogenous - in that case, OLS generally produces inconsistent estimates of *both* coefficients.

Now consider the following alternative set of assumptions

Assumption L2: $E[\epsilon_i z_i] = 0, E[z_i x_i] = 0$

Note the distinction between $(L1)$ and $(L2)$ arises in the second component - while $(L1)$ requires x_i to be uncorrelated with ϵ_i, $(L2)$ requires x_i to be uncorrelated with z_i.

One can easily show that (again assuming regularity and rank conditions hold) that $(L2)$ ensures identification of the causal effect β_2. To see this, decompose ϵ_i into its linear projection on x_i and a residual, i.e.

$$\epsilon_i = \lambda x_i + \widetilde{\epsilon}_i$$

and consider the transformed model

$$y_i = \widetilde{\beta}_1 x_i + \beta_2 p_i + \widetilde{\epsilon}_i \tag{2}$$

where $\widetilde{\beta}_1 = \beta_1 + \lambda$.

By construction,

$$E[\widetilde{\epsilon}_i x_i] = 0 \tag{3}$$

In addition,

$$E[\widetilde{\epsilon}_i z_i] = E[(\epsilon_i - \lambda x_i) z_i] = E[\epsilon_i z_i] - \lambda E[x_i z_i] = 0 \tag{4}$$

by $(L2)$. Together, (3), (4) imply that the transformed model (2) satisfies $(L1)$. Hence, applying IV to this model produces consistent estimates of $\widetilde{\beta}_1$ and β_2. While $\widetilde{\beta}_1 = \beta_1 + \lambda$ is not the causal effect of x_i on y_i, β_2 is the causal effect of p_i on y_i, so IV under $(L2)$ consistently estimates the parameter we are interested in.

There are a couple of intuitive ways to think about this result. First, for some intuition behind why this works, note that under $(L2)$, we could simply ignore x_i - lumping it in with the error term. This results in the model

$$y_i = \beta_2 p_i + (\beta_1 x_i + \epsilon_i)$$

Since z_i is uncorrelated with both x_i and ϵ_i, it is uncorrelated with the composite error term $(\beta_1 x_i + \epsilon_i)$. Hence, IV consistently estimates β_2. Of course, one would never do this in practice, as the resulting estimator would be inefficient relative to the one including x_i as a covariate. A second source of intuition behind the result is that because x_i and z_i are uncorrelated, the "transmitted bias" on β_2 described above disappears. This is again analagous to the more well-known OLS result - suppose that p_i is exogeous and x_i is endogenous - in this case OLS *can* consistently estimate the causal effect of p_i when p_i and x_i are uncorrelated. However, in a moment we argue that this is a much more powerful result in an IV setting.

In summary, we can obtain consistent estimates of the causal effect of p_i on y_i even if other covariates x_i are endogenous and we have no outside instruments for them. We feel that this is an underappreciated result for a number of reasons. First, it is always preferable to have more possible identifying assumptions - in some cases, one simply might be more willing to make assumption $(L2)$ than assumption $(L1)$. Second, an important distinction between $(L1)$ and $(L2)$ is that while $(L1)$ is not a directly testable set of assumptions[1], part of $(L2)$ is directly testable. Specifically, one can fairly easily check whether $E[z_i x_i] = 0$ in one's dataset. This can then allow one to relax the assumption that $E[\epsilon_i x_i] = 0$. Thirdly, taking somewhat of a Bayesian perspective, we feel that in some cases, verifying that $E[z_i x_i] = 0$ may make us more confident in the untestable assumption that $E[\epsilon_i z_i] = 0$. The basic idea here is that if ϵ_i is analagous to x_i (except for the fact that ϵ_i is unobserved to the econometrician), e.g. x_i are observed product characteristics, ϵ_i are unobserved (to the econometrician) product characteristics, a finding that $E[z_i x_i] \neq 0$ might make one worried that $E[\epsilon_i z_i] = 0$. In our empirical model we investigate this idea further.

Lastly, compare this result to the OLS result described above where p_i is exogeous and x_i is endogenous. In the OLS case, p_i and x_i will either be correlated or not - there is not much one can do to estimate the causal effect of p_i if they are correlated. On the other hand, in the IV case, there is the possibility that one has multiple instruments for p_i. In this case, one can explicitly look for potential instruments that satisfy $E[z_i x_i] = 0$. If one can find such an instrument (or instruments), we have shown that one can estimate β_2 consistently even with an endogenous x_i. This result is therefore important for the instrument selection issue when one is concerned about an endogenous x_i. More specifically, one can look for instruments that satisfy this property. Later, this ends up being a key goal of our empirical model. Even if one is reasonably comfortable assuming that x_i is exogenous, it seems to us that examining $E[z_i x_i]$ might be useful to examine possible "robustness" to violations of this assumption.

[1] It could be indirectly testable in the case where one has overidentifying restrictions, but those tests rely on auxiliary assumptions.

2.2 Non-linear Models

We next examine if this result holds up as we move to more flexible, non-parametric models. As an example, we consider the non-parametric IV model of Chernozhukov, Imbens, and Newey (2006) (CIN), i.e.

$$y_i = g(x_i, p_i, \epsilon_i) \tag{5}$$

where x_i and p_i are defined as above. Two important restrictions of the CIN model are that ϵ_i is a scalar unobservable and that g is strictly monotonic in ϵ_i. While this does allow for some forms of unobservable heterogeneous treatment effects (where the effect of p_i on y_i depends on unobservables) it is not completely flexible in this dimension. On the other hand, the model is completely flexible in allowing heterogenous treatment effects that depend on the observed covariates x_i. CIN normalize the distribution of ϵ_i to be $U(0,1)$ - this is WLOG because of the non-parametric treatment of g - intuitively, an appropriate g can turn the uniform random variable into whatever distribution one wants.

The analogue of causal effects in the CIN model are "quantile treatment effects". Specifically,

$$g(x_i', p_i', q_\tau) - g(x_i, p_i, q_\tau)$$

is the causal effect on y_i from moving from (x_i, p_i) to (x_i', p_i'), evaluated at the τth quantile of the ϵ_i distribution. Given the above normalization of ϵ_i to be $U(0,1)$, this is also the causal effect of moving from (x_i, p_i) to (x_i', p_i') conditional on $\epsilon_i = q_\tau$. As in the above linear model, we assume that we are only interested in estimating the causal effects of changing p_i. In other words, the "quantile treatment effects" we are interested in are all given a fixed x_i (i.e. involve $x_i' = x_i$).

Again ignoring regularity and rank conditions, the key identification assumption of CIN is

Assumption N1: (x_i, z_i) are jointly independent of ϵ_i

This independence condition is considerably stronger than the zero correlation conditions in the linear model, but that is what is typically required for non-parametric identification of these sorts of models. More importantly for our purposes, while this assumption allows arbitrary correlation between p_i and ϵ_i, it assumes that x_i is exogenous.

Our question is whether, as was done in the linear model, we can replace the assumption that ϵ_i is independent of x_i with an alternative assumption relying more on assumptions regarding the relationship between x_i and z_i. It turns out we can. Consider

Assumption N2: (x_i, ϵ_i) are jointly independent of z_i

To consider estimation under $(N2)$, we first show that $(N2)$ implies that $\epsilon_i$and $z_i$are independent *conditional* on x_i. To prove this, note that

$$\begin{aligned} p(z_i, \epsilon_i | x_i) &= \frac{p(z_i, \epsilon_i, x_i)}{p(x_i)} \\ &= \frac{p(z_i) p(\epsilon_i, x_i)}{p(x_i)} \\ &= p(z_i) p(\epsilon_i | x_i) \\ &= p(z_i | x_i) p(\epsilon_i | x_i) \end{aligned}$$

where the second and last equalities follow from $(N2)$.

What is the meaning of this result? $(N2)$ states directly that our instrument is valid (in the sense of being independent of ϵ_i) in the entire population. This simple implication of $(N2)$ says that our instruments *continue* to be valid even after conditioning on x_i. That is to say, conditioning on x_i does not generate correlations between z_i and ϵ_i. The importance of this result is quite intuitive - it says we can simply condition on x_i to avoid the problem of x_i being correlated with ϵ_i - doing this conditioning does not destroy the properties of our instrument.

To formally do this conditioning, assume that x_i has a discrete support to avoid technical issues. Pulling the x_i dependence into the g function, we get

$$y_i = g_{x_i}(p_i, \epsilon_i) \tag{6}$$

Think about estimating this transformed model separately for each possible value in the support of x_i. As just shown, conditional on being at each of these support points, ϵ_i and z_i are independent. Of course, because of the correlation between ϵ_i and x_i, the distribution of ϵ_i will vary across these support points. At each support point, renormalize the distribution of ϵ_i to be $U(0, 1)$ - this only involves changing g_i. This transformed model now satisfies $(N1)$ - hence, the CIN result suggests that we can estimate quantile treatment effects of this transformed model.

Importantly, because we have completely conditioned on x_i, our quantile treatment effects are conditioned completely on x_i. That is,

$$g_{x_i}(p'_i, q_\tau) - g_{x_i}(p_i, q_\tau)$$

is the causal effect on y_i from moving from (p_i) to (p'_i), evaluated at the τth quantile of the ϵ_i distribution *conditional* on x_i. These are different than the quantile treatment effects of the untransformed model (which would estimate the causal effect on y_i from moving from (p_i) to (p'_i), evaluated at the τth quantile of the *unconditional* ϵ_i distribution), but fine for many empirical purposes, particularly in empirical Industrial Organization .

Summarizing, we have shown that as in the linear model, we do not have to necessarily assume that the covariates x_i are exogenous to estimate the causal effect of p_i on y_i. We can instead look for instruments for p_i that appear to be independent of the covariates x_i. Note that assumption $(N2)$ is not quite as testable as in the linear case. Not only does z_i have to be independent of each of x_i and ϵ_i individually, but z_i has do be independent of the entire joint distribution of (x_i, ϵ_i). The only part of this that is directly testable is that z_i is independent of x_i. However, this still should be a useful test. In addition, again appealing to a pseudo Bayesian perspective, finding evidence that z_i is independent of x_i may be supportive of the assumption that z_i is independent of ϵ_i and the joint distribution (x_i, ϵ_i).

Before continuing, note that there is a third possible identifying assumption that one could also use to identify the above model. One could *directly* make the assumption that ϵ_i and z_i are independent *conditional* on x_i, i.e.

Assumption N3: (z_i, ϵ_i) are independent conditional on x_i

Identification of conditional quantile treatment effects under this assumption follows directly from the above. Note that while $(N2)$ implies $(N3)$, the reverse is not so. We think there are at least two important examples when this is the case. First, note that under $(N3)$, there can actually be correlation not only between z_i and x_i, but also between x_i and ϵ_i. Suppose, for example

$$\begin{aligned} z_i &= f^1(x_i) + \eta_i^1 \\ \epsilon_i &= f^2(x_i) + \eta_i^2 \end{aligned}$$

If η_i^1 and η_i^2 are independent (conditional on x_i), then $(N3)$ will hold, even though both z_i and ϵ_i are correlated with x_i. Given the structure of these two equations, this type of assumption might be appropriate when x_i's are can be thought of as being determined outside the economic model under consideration.

As a second example, suppose that z_i satisfies $(N2)$, i.e. (x_i, ϵ_i) are jointly independent of z_i. But suppose that the econometrician does not directly observe the instrument z_i. Suppose instead that what is observed is some function of z_i and x_i, i.e.:

$$z_i^* = h(z_i, x_i)$$

In this case, while the observed instrument z_i^* certainly does not satisfy $(N2)$, it does satisfy $(N3)$. Hence, the causal effect of the endogenous p_i will be identified. Note that this would also be the case if other random variables η_i that are independent of x_i and ϵ_i also entered the above equation, e.g.

$$z_i^* = h(z_i, x_i, \eta_i)$$

2.3 Combining Identification Assumptions

Note that one can use different types of the above identification assumptions for different covariates. For example, suppose we expand our demand model to the following

$$y_i = g(m_i, x_i, p_i, \epsilon_i) \tag{7}$$

where now both m_i and x_i are covariates. Again, suppose that we are only interested in estimating the causal effect of p_i on y_i. Consider the following assumption

Assumption N4: (x_i, ϵ_i) are jointly independent of z_i, conditional on m_i

Assumption $(N4)$ essentially combines assumption $(N2)$ on the x_i covariates and assumption $(N3)$ on the m_i covariates. To verify that we can identify conditional (on m_i and x_i) quantile treatment effects in this model, we just need to show that $(N4)$ implies that (z_i, ϵ_i) are independent conditional on x_i and m_i, i.e.

$$\begin{aligned} p(z_i, \epsilon_i | x_i, m_i) &= \frac{p(z_i, \epsilon_i, x_i | m_i)}{p(x_i | m_i)} \\ &= \frac{p(z_i | m_i) p(\epsilon_i, x_i | m_i)}{p(x_i | m_i)} \\ &= p(z_i | m_i) p(\epsilon_i | x_i, m_i) \\ &= p(z_i | x_i, m_i) p(\epsilon_i | x_i, m_i) \end{aligned}$$

Given this result, it follows from the above (treating $x_i = (x_i, m_i)$) that we can identify the conditional quantile treatment effects.

Why might we want to treat our covariates asymetrically? Recall our demand example. Suppose that m_i are market characteristics (e.g. the distribution of income, population density, etc.) and that x_i and ϵ_i are respectively, observed and unobserved (to the econometrician) product characteristics. Recall that p_i is price, z_i is an instrument for price, and y_i is demand for the product. If z_i, are e.g. input price shocks, it seems presumptous to assume that they are independent of general market characteristics. However, it does seem plausible that, conditional on market conditions, variation in z_i might be independent of product characteristics x_i and ϵ_i.

3 Bounding Bias

Moving back to the linear case, the above derivations consider the situation where the instrument is uncorrelated with the endogenous product characteristics. But what if one is unable to find such an instrument? It turns out that in these cases, one can often use the observed correlation between the instrument and the endogenous product characteristics to bound the possible bias on the price coefficient. For related ideas, see Nevo and Rosen (2006). Among other things, this can be used to choose between possible instruments.

In this section, we currently consider only two explanatory variables, although we extend this below. We also start by just examining the OLS case - i.e. where one explanatory variable is endogenous, and the question is how much bias is imparted on the other coefficient. Later we move to the IV case we have been discussing above.

Consider the following model:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

where all variables have been demeaned. Suppose that x_1 is potentially correlated with the residual ϵ, but x_2 is uncorrelated with ϵ. Our primary concern is to estimate the parameter β_2. Consider the OLS estimator formed by regressing y on x_1 and x_2.

$$\beta_{OLS} = (X'X)^{-1}X'y$$

where

$$X = \begin{bmatrix} x_{11} & x_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ x_{1N} & x_{2N} \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_N \end{bmatrix}$$

Substituting in, we get:

$$\begin{aligned} \beta_{OLS} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= \beta + (X'X)^{-1}X'\epsilon \end{aligned}$$

The second term is a bias term. Looking at the plim of this bias term in more detail, we have:

$$p\lim(X'X)^{-1}X'\epsilon = \begin{bmatrix} p\lim \frac{1}{N}\sum_i x_{1i}^2 & p\lim \frac{1}{N}\sum_i x_{1i}x_{2i} \\ p\lim \frac{1}{N}\sum_i x_{1i}x_{2i} & p\lim \frac{1}{N}\sum_i x_{2i}^2 \end{bmatrix}^{-1} \begin{bmatrix} p\lim \frac{1}{N}\sum_i x_{1i}\epsilon_i \\ 0 \end{bmatrix}$$

The zero in the second element of $X'\epsilon$ follows because of the assumption that x_2 is uncorrelated with ϵ. WLOG, normalize the variance of each of x_{1i} and x_{2i} to unity. This generates a bias term of

$$(X'X)^{-1}X'\epsilon = \begin{bmatrix} 1 & Cov(x_{1i}, x_{2i}) \\ Cov(x_{1i}, x_{2i}) & 1 \end{bmatrix}^{-1} \begin{bmatrix} Cov(x_{1i}, \epsilon_i) \\ 0 \end{bmatrix}$$

Inverting the matrix manually generates a bias vector of:

$$(X'X)^{-1}X'\epsilon = \begin{bmatrix} \frac{1}{1-Cov(x_{1i},x_{2i})^2}Cov(x_{1i},\epsilon_i) \\ \frac{-Cov(x_{1i},x_{2i})}{1-Cov(x_{1i},x_{2i})^2}Cov(x_{1i},\epsilon_i) \end{bmatrix}$$

We are only concerned with the second term in this bias vector, i.e.

$$bias = \frac{-Cov(x_{1i},x_{2i})}{1-Cov(x_{1i},x_{2i})^2}Cov(x_{1i},\epsilon_i)$$

The absolute value of this bias is

$$abs(bias) = \frac{abs(Cov(x_{1i},x_{2i}))}{1-Cov(x_{1i},x_{2i})^2}abs(Cov(x_{1i},\epsilon_i))$$

First note that this bias term is increasing in the absolute value of $Cov(x_{1i},x_{2i})$ over its feasible range ($-1 < Cov(x_{1i},x_{2i}) < 1$). This means that given any level of correlation between x_{1i} and ϵ_i, lower (absolute) values of $Cov(x_{1i},x_{2i})$ indicate lower values of bias.

Next, note that $Cov(x_{1i},x_{2i})$ is observed by the econometrician. Given this, our question is whether we can bound this bias. Unfortunately, $Cov(x_{1i},\epsilon_i)$ is not observed by the econometrician, and can in general can take any value from -∞ to ∞ (as long as $Var(\epsilon_i)$ is set high enough). Hence, we need to make some additional assumptions in order to bound this bias term. There are a couple of ways to proceed.

First, one could make a direct assumption on the possible range of $Cov(x_{1i},\epsilon_i)$. However, this seems like a strange term to be making a-priori assumptions on. Alternatively, note that the covariance of two variables is bounded by the product of their two variances, i.e.

$$\begin{aligned} abs(Cov(x_{1i},\epsilon_i)) &< SD(x_{1i})SD(\epsilon_i) \\ &< SD(\epsilon_i) \end{aligned}$$

This implies that that:

$$abs(bias) < \frac{abs(Cov(x_{1i},x_{2i}))}{1-Cov(x_{1i},x_{2i})^2}SD(\epsilon_i)$$

This bound can potentially be pretty tight. Suppose for example that x_{1i}, x_{2i}, and ϵ_i all contribute "equally" (in a causal sense) to y_i. This would be the case if we set $\beta_1 = 1$, $\beta_2 = 1$, and $SD(\epsilon_i) = 1$. Then if, for example, $Cov(x_{1i},x_{2i}) = 0.2$, the *maximal* bias is 0.2, or 20% - this maximum occurs when x_{1i} and ϵ_i are perfectly correlated.

It turns out that one can actually shrink these bounds a bit more. The reason is that if x_{1i} and x_{2i} are correlated and x_{2i} and ϵ_i are uncorrelated, then x_{1i} and ϵ_i cannot be perfectly correlated. However, this does not increase the bound by much when $Cov(x_{1i},x_{2i})$ is small, so we ignore this approach for now.

Of course, the above assumption that $SD(\epsilon_i) <= 1$ is one that could certainly seem arbitrary. Is there any natural upper bound for $SD(\epsilon_i)$? One somewhat natural bound might be the standard deviation of the dependent variable $SD(y_i)$. It is not necessarily the case that $SD(\epsilon_i)$ is less than $SD(y_i)$. However, there is a more primitive assumption that generates this result - that ϵ_i is positively correlated with $\beta_1 x_{1i} + \beta_2 x_{2i}$. This condition can also hold if ϵ_i is negatively correlated with $\beta_1 x_{1i} + \beta_2 x_{2i}$, but it cannot be too negatively correlated. Formally,

$$Var(y_i) = Var(\beta x_i) + Var(\epsilon_i) + 2Cov(\beta x_i, \epsilon_i)$$

Therefore:

$$\begin{aligned} Var(y_i) &> Var(\epsilon_i) \Leftrightarrow Var(\beta x_i) + 2Cov(\beta x_i, \epsilon_i) > 0 \\ &\Leftrightarrow Var(\beta x_i) + 2Corr(\beta x_i, \epsilon_i) SD(\beta x_i) SD(\epsilon_i) > 0 \end{aligned}$$

This clearly indicates that if $Corr(\beta x_i, \epsilon_i) > 0$, then $SD(\epsilon_i) < SD(y_i)$. But even if $Corr(\beta x_i, \epsilon_i) < 0$, then the condition will still hold unless $Corr(\beta x_i, \epsilon_i)$ is very negative and $SD(\epsilon_i)$ is reasonably high. For example, note that if we assume that the observed characteristic are "twice as important" as unobserved characteristics (in the sense that $SD(\beta x_i) > 2SD(\epsilon_i)$), then the condition must hold, even if $Corr(\beta x_i, \epsilon_i) = -1$.

A couple of more notes - in the BLP context, at least the price component of βx_i will be negatively correlated with ϵ_i. This is slightly problematic for the potential argument that $Corr(\beta x_i, \epsilon_i) > 0$ (but not for the potential argument that $SD(\beta x_i) > 2SD(\epsilon_i)$). Another way motivate this condition is using a hypothetical thought experiment. Suppose, we took a dataset (i.e. observed x_i's and y_i's) and forced all x_i's to their means. The question is what is the variance of the new y_i's. If one is willing to assume that the new y_i's are not as varied as the original y_i's then $SD(\epsilon_i)$ must be $< SD(y_i)$.

4 Application to the Discrete Choice Demand Literature

The above results can be straightforwardly applied to the literature on estimating demand systems in differentiated product markets. This literature often works with data across markets that is aggregated to the product level. For example, for each product j in market t, one typically observes p_{jt} (prices), X_{jt} (a vector of product characteristics), and s_{jt} (product j's market share in market t)

The logit-based utility model generates an estimating equation of the following form (see Berry (1994))

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = X_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

where ξ_{jt} is the econometric error (these are typically interpreted as either unobserved product characteristics or shocks to demand). Estimation of these models is typically done using IV/2SLS using instruments for p_{jt}. In other words, estimation proceeds allowing p_{jt} to be endogenous and correlated with ξ_{jt}, but under the assumption that product characteristics are exogenous, i.e. uncorrelated with ξ_{jt}. Instruments z_{jt} come from various sources, including, e.g., traditional cost shifters, characteristics of competing products in the market (see Bresnahan (1987) and BLP (1995) for motivation), and prices of the same product in other markets (see Hausman (1997) and Nevo (2001) for motivation).

Under the maintained assumption that the instruments are uncorrelated with ξ_{jt}, one will obtain consistent estimates of β and α as long as the product characteristics X_{jt} are uncorrelated with ξ_{jt}. However, if these product characteristics are correlated with ξ_{jt}, then estimates of *both* β and α will be biased - β suffers a direct bias due to the endogeneity of X_{jt}, and α generally suffers a transmitted bias. The derivations above, however, show that in the case where the instrument z_{jt} is uncorrelated with X_{jt}, there is *no* transmitted bias. In this case, α will be estimated consistently with the IV/2SLS estimator (though the estimate of β will still be biased).

There are at least a couple of approaches to using this result. First, one can simply look for instruments that satisfy this condition. Since the condition can be directly tested, this is relatively straightforward. Second, one can use theory to think about what choices of instruments might satisfy this condition. One interesting example of this uses cost shifters as instruments and is based on timing assumptions. Suppose that product characteristics X_{jt} are chosen before price p_{jt}. This seems like a natural assumption for many products where prices can be varied quite rapidly, but product characteristics need to be designed and planned in advance. In such a situation, one would want to use cost shocks that are realized by firms between the points in time . Such instruments would be uncorrelated with X_{jt} but likely correlated with p_{jt}. A specific example would be a situation in which there are futures markets for the price of an input. In this case, one would want to use the difference between the realized input price at the time when p_{jt} is set, and the expected future value of that input price at the time when X_{jt} was set. This by construction should be uncorrelated with X_{jt} . Of course, this correlation can be tested, which would be an implicit test of the timing assumption used to generate the condition.

One important additional point is that under these conditions, we have shown that α can be consistently estimated. α measures the marginal effect of price on utility. However, in many cases, a key goal of IO researchers is to estimate own and cross price elasticities/derivatives, which potentially depend on all parts of the model, i.e. both β and α. It is not directly obvious that these elasticities/derivatives can be estimated, since we cannot consistently estimate β. However, one can easily show that in this model, own and cross price derivatives can be written as:

$$\begin{aligned}
\frac{\partial s_{jt}}{\partial p_{jt}} &= -\alpha s_{jt}(1 - s_{jt}) \\
\frac{\partial s_{jt}}{\partial p_{kt}} &= \alpha s_{jt} s_{kt}
\end{aligned}$$

This indicates that price derivatives (and elasticities) can be written *only* as a function of the data and the estimated α - *not* of the estimated β. Hence these price derivatives and elasticities can be estimated consistently. Of course, note that we cannot recover elasticities w.r.t. characteristics, since those will obviously depend on β.

This result also applies to more flexible differentiated product demand models. The estimating equation in the nested logit model can be expressed as:

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = X_{jt}\beta - \alpha p_{jt} + \sigma \ln(s_{jt|g}) + \xi_{jt}$$

where $s_{jt|g}$ measures product j's within-group (nest) market share, and σ is a parameter measuring how substitution patterns differ within and across nests. Since $s_{jt|g}$ is by construction endogenous in these models, one needs at least two instruments to deal with the endogenous p_{jt} and $s_{j|g}$. Our results above indicate that if these instruments are uncorrelated with X_{jt}, then IV/2SLS estimation of the above equation produces consistent estimates of α and σ, even if X_{jt} is endogenous. One can easily show that in these models, price derivatives can be written as:

$$\begin{aligned}
\frac{\partial s_{jt}}{\partial p_{jt}} &= -\alpha s_{jt}\left(\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} s_{jt|g} - s_{jt}\right) \\
\frac{\partial s_{jt}}{\partial p_{kt}} &= \begin{cases} \alpha s_{kt}\left(\frac{\sigma}{1-\sigma} s_{jt|g} + s_{jt}\right) & \text{if } j \text{ and } k \text{ in same nest} \\ \alpha s_{jt} s_{kt} & \text{otherwise} \end{cases}
\end{aligned}$$

Again, since these do not depend on β, these can be consistently estimated as well.

Lastly, consider the random coefficients model. The estimating equation in this model can be expressed as

$$\delta(\{X_{jt}, p_{jt}, s_{jt}\}_{j=1}^{J_t}; \Sigma) = X_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

where Σ are parameters capturing the distribution of the random coefficients. Since Σ enters non-linearly, one cannot use IV/2SLS here, and one typically proceeds by using GMM. Computation of $\delta\,\{X_{jt}, p_{jt}, s_{jt}\}_{j=1}^{J_t}; \Sigma)$ typically requires simulation and a contraction mapping. More instruments are needed in these models, as one needs additional instruments to identify the parameters Σ. However, one can show that as long as these instruments are uncorrelated with X_{jt}, one can obtain consistent estimates of α and Σ, even if X_{jt} is incorrectly assumed exogenous. Moreover, like in the logit and nested logit models, one can show (using the Berry (1994) inversion) that one can calculate price elasticities/derivatives without knowledge of β.

5 Empirical Example

We demonstrate the ideas developed in this paper in an empirical example using data from the cable television industry. The data report the number of offered Basic and Expanded Basic cable services, and the prices, market shares, and number of cable programming networks offered on each service for a sample of 4,447 cable systems across the United States.[2]

Summary statistics for each of the variables follow the appendix. We consider a simple example based as closely as possible on the theory described above. We estimate a logit demand system for each of the products offered by the cable system in each market. The key explanatory variables are price (tp) and number of offered cable programming networks (tx).

We consider a number of instruments for price, all based on variables that influence the marginal cost of providing cable service. The primary marginal cost for cable systems are "affiliate fees", per-subscriber fees that they must pay to television networks (e.g. ESPN) for the right to carry that network on their cable system. These instruments are:

1) Homes Passed (hp).- If larger cable systems have better bargaining positions with content providers, they may receive lower affiliate fees.2)

2) Franchise Fee (franfee).- Franchise fees are payments made by cable systems to the local governing body in return for access to city streets to install their cable systems. Systems facing higher franchise fees may have higher marginal costs and therefore charge higher prices. This was the primary price instrument used in Goolsbee and Petrin (2005).

3) Average Affiliate Fees (tcx).- Kagan Media collects information about the average (across systems) affiliate fee charged for the vast majority of television networks offered on cable. This variable calculates the average fee for the networks offered by each cable system in the sample.

4) MSO Subscribers (msosubs) -. Multiple System Operators, or MSOs, are companies that own and operate multiple cable systems across the country (e.g. Comcast, Cox). This variable proxies for bargaining power of cable systems in (nationwide) negotiations with television networks.

5) Prices in other markets (tip, tipst, tipreg).- MSOs generally negotiate the affiliate fees they will pay to television networks on behalf of all the systems in the corporate family. As such, the marginal cost for providing cable service should be similar for cable systems within an MSO. If demand shocks are uncorrelated across these systems, cable prices in other markets for systems within the same MSO might be a good instrument for prices in any given market. Hausman (1998) and Nevo (2001) have used the this strategy of finding instruments in the cereal market and Crawford (bundling paper) has used it in cable markets. Because it relies heavily on the lack of correlation in demand errors across markets, we construct three measures of this instrument:

[2] The data have a lot more, esp. the identity of offered networks for each bundle, demographic info in each market, etc.

the average price for each offered cable service within an MSO excluding the current system (tip), the average price for each offered service within an MSO excluding those systems in the current systems state (tipst), and the average price for each offered service within an MSO excluding those systems in the current systems' census region (tipreg).[3]

Here are the preliminary results. First we show the results of the first-stage regression of price (tp) on all the explanatory variables and the instruments (we do this regression separately for each instrument, one at a time).[4] Most of the results are of the correct sign and of reasonable magnitude.

Variable	fols	fivhp	fivfr~e	fivtcx	fivms~s	fivtip	fivti~t	fivti~g
tx	0.000	0.256	0.247	0.016	0.254	0.240	0.254	0.249
	0.000	0.009	0.009	0.021	0.009	0.009	0.010	0.010
tp	1.000							
	0.000							
hp		-0.051						
		0.006						
franfee			-0.273					
			0.047					
tcx				1.287				
				0.110				
msosubs					-0.337			
					0.019			
tip						0.642		
						0.017		
tipst							0.487	
							0.020	
tipreg								0.454
								0.020

Next are the associated IV results using each instrument as the single instrument for price (standard errors below the estimates), treating product characteristics as exogenous. Also reported is the estimated impact to mean utility of an additional cable programming network. As expected,

[3] We use the four major Census regions: NE, S, MW, and W.

[4] Other variables not reported in the table are dummy variables for goods 1, 2, 3, and 4 in each market as well as dummy variables indicating for each good whether the next higher goods are also offered, i.e. ind31 = a dummy variable in demand for good 1 indicating whether good 3 was also offered in that market.

instrumenting for price generally yields a larger estimated price sensitivity than that generated by OLS.

Variable	ols	ivhp	ivfra~e	ivtcx	ivmso~s	ivtip	ivtipst	ivtip~g
tp	-0.038	-0.022	-0.048	-0.024	-0.025	-0.070	-0.090	-0.078
	0.002	0.022	0.030	0.015	0.010	0.005	0.008	0.008
tx	0.029	0.025	0.032	0.026	0.026	0.038	0.042	0.039
	0.002	0.005	0.007	0.004	0.003	0.002	0.003	0.003

Next, the above results are use to compute the associated average estimated own-price elasticity (averaged across all products).[5] As expected, the means basically mirror the estimated price sensitivities in the table above.

Variable	Obs	Mean	Std. Dev.	Min	Max
elastols	5807	-.4299353	.2106194	-2.730194	-.0107715
elastivhp	5807	-.2394281	.1100084	-1.507058	-.0075227
elastivfra~e	5807	-.5390264	.2733753	-3.458999	-.011781
elastivtcx	5807	-.267501	.1240874	-1.686831	-.0081377
elastivmso~s	5807	-.2776902	.1292619	-1.752155	-.0083485
elastivtip	5807	-.8115802	.4374062	-5.397118	-.013089
elastivtipst	5807	-1.072409	.6106484	-7.184474	-.0122947
elastivtip~g	5807	-.9204701	.5036151	-6.105941	-.0131605

Lastly, we consider whether each of these excluded instruments is correlated with the product characteristic (tx). All are statistically significant, except for the price of cable service at other systems within the same MSO (tip).

Variable	xols	xhp	xfran~e	xtcx	xmsos~s	xtip	xtipst	xtipreg
tp	0.428							
	0.017							
hp		0.157						

[5]Because cable services are cumulative, it is technically cleaner to look just at the highest-quality good offered in each market. Doing so yields qualitatively similar results.

```
             |                 0.008
      franfee |                           0.935
             |                           0.064
         tcx |                                     4.698
             |                                     0.030
     msosubs |                                               0.217
             |                                               0.027
         tip |                                                        -0.024
             |                                                         0.027
       tipst |                                                                  -0.213
             |                                                                   0.030
      tipreg |                                                                            -0.169
             |                                                                             0.029
-------------------------------------------------------------------------------------------------
                                                                                  legend: b/se
```

This suggest that the results using tip are likely to be the most reliable. We can formalize this comparison using our upper bounds. In particular suppose we use (**??**) to calculate upper bounds on the absolute value of the bias, imposing the assumption the assumption that $\frac{1}{N}X_1 M_W \epsilon < \sqrt{\frac{1}{N}X_1 M_W X_1}\sqrt{\frac{1}{N}y M_W y}$. we get the following bias bounds using the various instruments:

```
    biashp = 0.42032582
biasfranfee = 0.45246954
   biastcx =  2.4384633
biasmsosubs =  0.0829468
   biastip =  0.0047211
 biastipst = 0.05610395
biastipreg = 0.04769696
```

6 Conclusions

7 Appendix

Variable	Obs	Mean	Std. Dev.	Min	Max
origcid	5807	3196.54	1874.354	1	6481
year	5807	2002	0	2002	2002

nprod	5807	1.535905	.6809579	1	4
prod	5807	1.267952	.5175304	1	4
y98	5807	0	0	0	0
y99	5807	0	0	0	0
y00	5807	0	0	0	0
y01	5807	0	0	0	0
y02	5807	1	0	1	1
chancap	5807	46.1307	20.93914	4	150
hp	5807	5.204854	13.50131	.018	418.2
franfee	5807	1.480799	1.819802	0	8
msosystems	5807	551.5468	513.5846	0	1353
msosubs	5807	3.863265	4.528929	0	13.75
msopaysubs	5807	3096293	4116557	0	1.34e+07
logsrat	5807	-.0968512	1.507517	-6.587178	7.090035
s	5807	.4618122	.2740264	.0006661	.9972565
tp	5807	22.46918	8.197811	.95	80.85001
tip	5459	23.62627	6.728524	7.5	56.33334
tipst	4832	23.70755	6.706235	9.59	51.78333
tipreg	4548	23.5211	6.652446	10.95	51.06667
tind1	5807	1	0	1	1
tind2	5807	.2342001	.4235343	0	1
tind3	5807	.0316859	.1751776	0	1
tind4	5807	.0020665	.0454154	0	1
tind21	5807	.4367143	.4960215	0	1
tind31	5807	.0909247	.2875268	0	1
tind41	5807	.0082659	.0905482	0	1
tind32	5807	.0613053	.2399102	0	1
tind42	5807	.0061994	.0784987	0	1
tind43	5807	.0041329	.0641605	0	1
ind1	5807	.7657999	.4235343	0	1
ind2	5807	.2025142	.401908	0	1
ind3	5807	.0296194	.1695496	0	1

ind4	5807	.0020665	.0454154	0	1
ind21	5807	.2025142	.401908	0	1
ind31	5807	.0296194	.1695496	0	1
ind41	5807	.0020665	.0454154	0	1
ind32	5807	.0296194	.1695496	0	1
ind42	5807	.0020665	.0454154	0	1
ind43	5807	.0020665	.0454154	0	1
indtop	5807	.7657999	.4235343	0	1
tx	5807	17.44946	10.61949	0	64
tcx	5807	4.573625	2.422858	0	10.52

References

[1] Berry, S.T. "Estimating Discrete Choice Models of Product Differentiation." *RAND Journal of Economics*, Vol. 25 (1994), pp. 242-262.

[2] ———, and Pakes, A. "Estimating the Pure Hedonic Discrete Choice Model." Working Paper, Yale University, 1999.

[3] ———, Levinsohn, J.A., and Pakes, A. "Automobile Prices in Market Equilibrium." *Econometrica*, Vol. 63 (1995), pp. 841-890.

[4] Bresnahan, T.F. "Competition and Collusion in the American Auto Industry: The 1955 Price War." *Journal of Industrial Economics*, Vol. 35 (1987), pp. 457-482.

[5] ———, T.F., Stern, S., and Trajtenberg, M. "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s." *RAND Journal of Economics*, Vol. 28 (1997), pp. S17-S44.

[6] Chernozhukov, Imbens, and Newey (2006) "Instrumental Variable Estimation of Non-separable Models" *Journal of Econometrics*, 139, 1 4-14

[7] Crawford, G. "The Impact of the 1992 Cable Act on Household Demand and Welfare." *RAND Journal of Economics*, Vol. 31 (2000), pp. 422-449.

[8] ———, and Shum, M. (2006) "The Welfare Effects of Endogenous Quality Choice: The Case of Cable TV", mimeo, Johns Hopkins

[9] Fan, Y. (2009) "Market Structure and Product Quality in the US Daily Newspaper Market", mimeo, Yale

[10] Goolsbee and Petrin (2004) "The consumer gains from direct broadcast satellites and the competition with cable TV", *Econometrica*, 72, 2. pp. 351-381

[11] Hausman, J. (1997) "Valuation of new goods under perfect and imperfect competition" in *The Economics of New Goods*, NBER, edited by Tim Bresnahan and Robert Gordon

[12] Nevo, A. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, Vol. 69 (2001), pp. 307-342.

[13] Nevo, A. and Rosen, A. (2008) "Identification with Imperfect Instruments", NBER working paper

[14] Pioner, H. (2008) "Semiparametric Identification of Multidimensional Screening Models", mimeo, Chicago

[15] Sweeting, A (2008) "Dynamic Product Repositioning in Differentiated Product Industries: The Case of Format Switching in the Commercial Radio Industry", mimeo, Duke