Estimating Price Elasticities in Differentiated Product Demand Models with Endogenous Characteristics*

Daniel A. Ackerberg and Gregory S. Crawford

March, 2007

1 Introduction

1) It's hard to believe the standard assumption that product characteristics in demand models are exogenous.

2) Maybe even harder to find instruments for endogenous product characteristics - 1) probably need alot of them. 2) if part of the error term is an "unobserved product characteristic" that is also chosen by the firm, it is hard to imagine one could ever find an "instrument" that is correlated with the observed product characteristics but uncorrelated with the unobserved product characteristics.

3) Alternative approaches - Crawford Shum approach - explicitly model firms choice of characteristics. nice approach, but definitely some limitations. Gets very hard very quickly - CS can only look at monopoly situations with one, observed product characteristic. . Ask Greg about identification issues.

4) Ackerberg, Berry, Benkard, and Pakes briefly mention a possible approach to dealing with endogenous product characteristics. This approach is similar to that of Olley-Pakes and the literature stemming from there - basically one assumes that . Two potential issues with this method are: 1) can't deal with endogenous unobserved product characteristics, 2) relies on non-directly testable informational assumptions on unobservables. In contrast, our identification will hinge on at least partially testable informational assumptions on observables (our price instruments). In many cases, one will need to rely on informational assumptions on unobservables as that might

*Preliminary and incomplete: Please do not cite or quote. Comments welcome. We are grateful to Kei Hirano, Jin Hahn, and seminar participants at New York University and Northwestern University for helpful comments. Correspondence may be sent to Daniel A. Ackerberg, Bunche Hall 8379, Department of Economics, UCLA, Box 951477, Los Angeles, CA, 90095-1477, phone 310-794-5427, email ackerber@econ.ucla.edu or Gregory S. Crawford, Department of Economics, University of Arizona, Tucson, AZ 85721-0108, phone 520-621-5247, email crawford@eller.arizona.edu.

be the only alternative. But it seems preferable to make such an assumption on observables if possible .

5) Our paper starts with the observation that one does not always need to estimate causal effects of changing characteristics - for many questions, e.g. most short-run antitrust questions, one is primarily interested in own and cross price elasticities. For other questions, e.g. hedonic issues, causal effects of characteristics are likely more important.

6) If this is the case, we show that under some conditions, one can consistently estimate these price elasticities. These conditions concern correlations between instruments and the endogenous product characteristics. Note that we are not really doing anything different than standard, e.g. BLP model - only difference would be in thinking about what makes a good instrument, and maybe trying to construct instruments that satisfy these properties, possibly from existing instruments..

7) Sheds light on what make good price instruments in these demand models - in particular, what make good price instruments when product characteristics are endogenous.

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 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,

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

In addition,

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

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

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

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 $E\left[z_i x_i\right]=0$ in one's dataset. It seems to us that if this condition appears to hold, there is no reason to make the non-directly testable assumption that $E\left[\epsilon_i x_i\right]=0$. Thirdly, taking somewhat of a Bayesian perspective, we feel that in some cases, verifying that $E\left[z_i x_i\right]=0$ may make us more confident in the untestable assumption that $E\left[\epsilon_i z_i\right]=0$. The basic idea here is that if ϵ_i is analagous to x_i (except for the fact that ϵ_i is

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

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. It seems to us that one should be looking 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 considering $E[z_i x_i]$ might be useful to examine possible "robustness" to violations of this assumption.

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 ϵ_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 slightly different than the quantile treatment effects of the untransformed model, but fine for our empirical purposes.

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 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 beWe independent of product characteristics x_i and ϵ_i. The next section formulates an economic model of endogenous produxt characteristics and demand that can be used to assess when this might be the case.

3 An Empirical Model of Demand with Endogenous Product Characteristics

To summarize, the above econometric discussion suggests that if we can find instruments for price that are uncorrelated or independent of observed product characteristics, we can identify price derivatives and elasticities. This is true even if product characteristics are endogenous w.r.t. demand unobservables. We now turn to investigating how one might find or construct such instruments. To do this, we will consider a formal model where firms decide on product characteristics and prices.

An important aspect of the models we consider is that we will assume that firms choose product characteristic *before* they decide on prices. This assumes that price is in some sense a more flexible decision than product characteristics. Obviously the appropriability of such an assumption will depend on the product of study, but it does seem to us that across a wide range of industries, price is probably a more "variable" decision than are product characteristics. For example, for manufactured products, one typically needs to decide on product characteristics a considerable time before production even occurs. In any case, as we will see, this separation in time will be important for us to find instruments that satisfy the conditions we are looking for.

Another important characteristic of our models is that they explicitly consider two types of costs: production costs and development costs. Development costs represent the fixed costs of developing a product with a particular set of characteristics, production costs represent the marginal cost of producing/distributing a unit of a product with a particular set of characteristics. For simplicity, we will assume that there are constant returns to scale, i.e. production costs are constant in the number of units produced, but the model would not fundamentally change if this assumption were weakened.

In all cases we will think about a demand function of the form:

$$y_i = g(m_i, x_i, p_i, \epsilon_i) \tag{8}$$

where i indexes markets. Conceptually, it is easiest to think of having data across markets, where in each market there is exactly one differentiated product for sale. In this case, m_i represent market conditions in market i, x_i is a vector of the observed characteristics of the product for sale in market i, p_i is the price in market i, and y_i are sales of the product. However, one can interpret all of what we do in a multiproduct, potentially oligopolistic, setting. In this case, y_i and p_i are vectors of quantities and prices of all the products in the market, and x_i is a matrix of the sets of product characteristics across products.

A distinction we will focus on is the interpretation of ϵ_i. Recall that there are at least two interpretations of ϵ_i. On one hand, ϵ_i could represent the unobserved analogue of x_i, i.e. unobserved (to the econometrician) product characteristics. Another possibility is that ϵ_i represent unobserved (to the econometrician) market variables, i.e. the analogue of m_i. We actually discuss three separate models. The first interprets ϵ_i as representing unobserved market variables, e.g. demand shifters. In this model, there are no product characteristics that are unobserved by the econometrician. The second model interprets ϵ_i as an unobserved product characteristic - there are no unobserved market variables that influence demand. The third model is most general, interpreting ϵ_i as a combination of both types of unobservables. The assumptions we need necessary to identify price effects will differ across these models.

Lastly, note that we proceed without making functional form assumptions on the various equations. Because of this we generally assume complete independence of unobserved shocks affecting various decisions. Talk about how in some cases this may actually be WLOG. With specific functional forms, we could probably relax some of these independence assumptions to mean independence or uncorrelatedness assumptions.

3.1 Unobserved Market Conditions

Define μ_i as a set of market variables that are unobserved by the econometrician. This contrasts with m_i, which are observed market variables. In the current section, we consider a model where the demand side unobservable ϵ_i is determined by these unobserved market variables, i.e.

$$\epsilon_i = h(\mu_i)$$

Note of course that all the elements of μ_i do not necessarily impact ϵ_i. Some components of μ_i may only be directly relevant for the supply side of the model.

We first model the costs of firms. Again, we consider two distinct types of costs, development costs and production costs. Denote the cost of developing a product with characteristics x_i as

$$c^d(x_i; m_i, \mu_i, z_i^d) \tag{9}$$

These development costs may depend on both observed and unobserved market conditions (m_i, μ_i). We also introduce a new set of variables z_i^d - these represent supply side shocks that impact the cost of development.

Similarly, suppose the cost of producing a product with characteristics x_i is

$$c^p(x_i; m_i, \mu_i, z_i^p)$$

Again, these production costs depend on observed and unobserved market conditions, along with supply side shocks that impact the cost of production.

Regarding timing in the model, suppose that firms make decisions regarding product characteristics at time t^d, prior to producing and selling the product at time t. Consider a firm deciding on its product characteristics x_i at time t^d. Given the above formulation of development costs, these decisions will clearly depend on m_i, μ_i, and development cost shocks z_i^d. Note that the dependence on m_i and μ_i comes from two sources. First, they both affect the costs of development in (2), second, the both affect demand in (8).

In addition, a forward looking firm will also want to consider the impact of its development decisions on production costs. As such, the firm will try to make inferences about production cost shocks z_i^p. To formalize these possible inferences, consider the firm's information set at time t

$$I_i^{t^d} = \{m_i, \mu_i, z_i^d, z_i^{p1}\}$$

There is a new variable in this information set, z_i^{p1}. This represents shocks that do not directly impact development costs, but that impact the firms percieved distribution of z_i^p. Note that z_i^{p1}, for example, could include particular elements of z_i^p, if that element happened to have been observed by the firm at or prior to time t_d. z_i^{p1} could also contain signals that are informative to the firm regarding on z_i^p. Given the above information set, the firm has perceptions of the distribution of z_i^p, i.e.

$$f(z_i^p | I_i^{t^d})$$

This percieved distibution of z_i^p is sufficient to examine the firm's choice of product characteristics x_i. More specifically, in this model, the firm's choice of characteristics will be of the form:

$$x_i = f_x(I_i^{t^d}) = f_x(m_i, \mu_i, z_i^d, z_i^{p1})$$

Again, we assume that there is a time period between when product characteristics are chosen and when price is chosen. In this time period, the actual production cost shocks z_i^p are realized. What we want to do now is decompose these production shocks into components that were predictable at time t^d (i.e. depend on the information set at time t^d) and components that are independent of this information set.

This decomposition is actually fairly easy to do. For example, suppose that z_i^p has only one element. Consider the cumulative distribution function of this z_i^p,

$$F(z_i^p | I_i^{t^d})$$

Inverting this CDF generates the equation

$$z_i^p = F^{-1}(z_i^{p2}|I_i^{t^d}) = h(z_i^{p2}, I_i^{t^d})$$

where the newly defined z_i^{p2} is independent of $I_i^{t^d}$ and has a uniform distribution $U(0,1)$. This equation decomposes z_i^p into two components - a part that is known at time t^d ($I_i^{t^d}$), and a part that is not known at time t^d (z_i^{p2}) and by construction independent of the information set at t^d. In other words, z_i^{p2} can be interpreted as the part of z_i^p that is not predictable given $I_i^{t^d}$. Note that it is possible that $z_i^p = h(z_i^{p2})$. This would be the case if z_i^p was independent of $I_i^{t^d}$ to start with.

This decomposition can also be done in the case where z_i^p is a vector of shocks - one simply can do these CDF inversions sequentially (see the appendix). This results in a vector of "innovations" z_i^{p2} that are independent of the information at t^d. To make things simpler however, we will assume the existence of a subset of variables in z_i^p that are independent of $I_i^{t^d}$. Call this subset z_i^{p2}.

Now lets move back to the firms decision of product characteristics at time t^d. Given the above informational structure, the firms optimal choice of product characteristics will take the form:

$$x_i = f_x(m_i, \mu_i, z_i^d, z_i^{p1})$$

Note that m_i, μ_i may affect After choosing these characteristics, the firm moves to time t, observing z_i^{p2} (and thus z_i^p). The firm now decides on prices

$$p_i = f_p(x_i, m_i, \mu_i, z_i^p)$$

.

Now to the pricing decision, at time t - think of the $(z_{i1}^{p2}, ..., z_{iK}^{p2})$ being realized.

Start by assuming that observed supply shocks are independent of

This decomposition of z_i^p into components actually allows us to think about things much more simply. Assume that one of these z is completely independent, in the sense that it is random something that is

$$\nu_i^{\backslash p} = (\nu_i^{p1}, \nu_i^{p2}) = (\nu_i^{p1}, \nu_i^{p2}(\nu_i^{p1}, m_i, \mu_i))$$

ν_i^{p1} are the elements of the production shocks that are observed to the firm when making their development decisions. In contrast, ν_i^{p2} are not observed

3.2 Unobserved Product Characteristics

We start with a model where we interpret ϵ_i as an unobserved product characteristic. We will start by considering ϵ_i symetrically to x_i. That is, just as the firm makes decisions on what x_i

they will have for their product, they decide on what ϵ_i they will have for their product. Later, we will consider asymetric models where ϵ_i is determined according to a different process than x_i.

4 Notes on Bounding Bias (from acnotes.tex)

First I look at 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 I move to the IV case we have been thinking about.

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

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

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 this bias term in more detail, we have:

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

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

[2] **gsc: Don't we need to be a little more careful here? Strictly speaking the second element is only zero asymptotically, so should we call this the asymptotic bias? And have plim's in front of all the 1/N terms?**

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 $Cov(x_{1i}, \epsilon_i)$. This seems like a strange term to be making assumptions on though.

Second, 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 $SD(\beta x_i) > 2SD(\epsilon_i)$).

Also, another way motivate this condition is using a hypothetical though 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 $< SD(y_i)$.

4.1 An Alternative Derivation of the above Bound

You might want to ignore this section - it's just a different way to end up with the same result as above - the reason it is here is that I wasn't completely sure of this result a-priori, and once I went through it, no sense in deleting it. This might also be an easier way to analyze the case when X_1 is more than single dimensional. Anyway....

Next consider the alternative formulation of the model

$$\begin{aligned} Y &= X_1\beta_1 + X_2\beta_2 + \epsilon \\ &= X_1\beta_1 + X_2\beta_2 + (P_{X_1} + I - P_{X_1})\epsilon \end{aligned}$$

where $P_{X_1} = X_1(X_1'X_1)^{-1}X_1'$. Define γ as the hypothetical regression coefficient if one regressed ϵ on X_1. Since $P_{X_1}\epsilon = X_1(X_1'X_1)^{-1}X_1'\epsilon = X_1\gamma$ we can rewrite the model as:

$$Y = X_1(\beta_1 + \gamma) + X_2\beta_2 + (I - P_{X_1})\epsilon$$

(NB: I'm not sure if we want γ to be the actual regression coefficient if one regressed ϵ on X_1 with the dataset at hand, or if we want it to represent the plim of this regression coefficient. In any case, I think the difference will asymptotically disappear). Again consider OLS estimation of this model

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

Substituting in, we get:

$$\begin{aligned} \beta_{OLS} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X_1(\beta_1 + \gamma) + X_2\beta_2 + (I - P_{X_1})\epsilon) \\ &= \begin{bmatrix} \beta_1 + \gamma \\ \beta_2 \end{bmatrix} + (X'X)^{-1}X'\epsilon^* \end{aligned}$$

where

$$\epsilon^* = (I - P_{X_1})\epsilon$$

The second term is again a bias term. With the same normalizations as before, we have:

$$(X'X)^{-1}X'\epsilon^* = \begin{bmatrix} 1 & Cov(x_{1i}, x_{2i}) \\ Cov(x_{1i}, x_{2i}) & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ X_2'\epsilon^* \end{bmatrix}$$

The zero in the first element of $X'\epsilon^*$ follows because $X_1'(I - P_{X_1}) = 0$. Inverting the matrix manually generates a bias vector of:

$$(X'X)^{-1}X'\epsilon = \begin{bmatrix} \frac{-Cov(x_{1i},x_{2i})}{1-Cov(x_{1i},x_{2i})^2} X_2'\epsilon^* \\ \frac{1}{1-Cov(x_{1i},x_{2i})^2} X_2'\epsilon^* \end{bmatrix}$$

The second term in this bias vector is:

$$
\begin{aligned}
bias &= \frac{1}{1-Cov(x_{1i},x_{2i})^2}X_2'\epsilon^* \\
&= \frac{1}{1-Cov(x_{1i},x_{2i})^2}X_2'(I-P_{X_1})\epsilon \\
&= -\frac{1}{1-Cov(x_{1i},x_{2i})^2}X_2'X_1(X_1'X_1)^{-1}X_1'\epsilon \\
&= \frac{-Cov(x_{1i},x_{2i})}{1-Cov(x_{1i},x_{2i})^2}Cov(x_{1i},\epsilon_i)
\end{aligned}
$$

the same result as before.

4.2 IV Situation

Same essential argument goes through for an IV estimator when x_{2i} is also endogenous. We assume the existance of an instrument z_i that is correlated with x_{2i} but uncorrelated with ϵ_i. Note that we will not be instrumenting for x_{2i}, i.e. our instrument matrix Z is equal to:

$$
Z = \begin{bmatrix} x_{11} & z_1 \\ . & . \\ . & . \\ x_{1N} & z_N \end{bmatrix}
$$

The IV estimator is given by:

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

This bias term is now:

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

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

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

Inverting the matrix manually generates a bias vector of:

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

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

$$bias = \frac{-Cov(z_i,x_{1i})}{Cov(z_i,x_{2i})-Cov(z_i,x_{1i})Cov(x_{1i},x_{2i})}Cov(x_{1i},\epsilon_i) \tag{10}$$

Note that this bias term is of the same magnitude as in the former case where x_{2i} was assumed endogenous. To see this, suppose that the instrument z_i generates half the variation in x_{2i}. Then $Cov(z_i,x_{2i}) = 0.5Var(x_{2i}) = 0.5$ and $Cov(z_i,x_{1i}) = 0.5Cov(x_{2i},x_{1i})$ (this second equation holds if the correlation between x_{2i} and x_{1i} is generated equally by the the z_i part of x_{2i} and the other part of x_{2i}). In this case, the 0.5's cancel out and we get the same expression as above.

I'm not convinced that the absolute value of this bias term is necessarily increasing in $Cov(z_i,x_{1i})$. As $Cov(z_i,x_{2i}) \longrightarrow 1$ it definitely does though. Regardless, however, this formula does seem to indicate that if one is choosing between instruments, one does not necessarily want to pick the instrument with the smallest $Cov(z_i,x_{1i})$ - the strength of the instrument, $Cov(z_i,x_{2i})$, is also relevant for the bias. In any case, again, all the elements of this bias term are estimable except for $Cov(x_{1i},\epsilon_i)$ (which doesn't depend on choice of instrument). Hence one could simply estimate the first term of the above for each instrument and pick the lowest.

As in the prior section, one can also bound the bias. The absolute value of the bias is given by:

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

which by the above

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

So if $SD(\epsilon_i)$ can be bounded as above, we have a bound on $abs(bias)$.

4.3 Bias with Additional Covariates

To this point we have considered a model with just one endogenous variable (x_2, e.g. price) and one possibly endogenous variable (x_1, e.g. a product characteristic). This subsection derives the bias formula when there are additional exogenous covariates.

Consider the following more general model model:

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

where W_i is a $R \times 1$ vector of additional exogenous covariates (e.g. market characteristics) and β_W measures their impact on demand for good i. Let $\tilde{X}$ and $\tilde{Z}$, the matrix of explanatory variables and instruments, now be given by

$$\tilde{X} = \begin{bmatrix} x_{11} & x_{21} & W_{11} & . & W_{R1} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{1N} & x_{2N} & W_{1N} & . & W_{RN} \end{bmatrix} = [X\ W] \qquad \tilde{Z} = \begin{bmatrix} x_{11} & z_1 & W_{11} & . & W_{R1} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{1N} & z_N & W_{1N} & . & W_{RN} \end{bmatrix} = [Z\ W]$$

where X and Z are as defined in the previous section and W is the $N \times R$ matrix of additional exogenous variables.

The IV estimator is now :

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

where $\beta \equiv (\beta_1\ \beta_2\ \beta_W)'$ is the $(R+2) \times 1$ vector of parameters.

The bias term is now

$$(\tilde{Z}'\tilde{X})^{-1}\tilde{Z}'\epsilon = \begin{bmatrix} Z'X & Z'W \\ W'X & W'W \end{bmatrix}^{-1} \begin{bmatrix} Z'\epsilon \\ W'\epsilon \end{bmatrix}$$

As earlier, we are particularly interested in the bias on β_2, but now have to allow for the additional influence of W on that bias.

To calculate the bias in the presence of W, we rely on the formula for partitioned regression (e.g. Greene, p.33):

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} F_1 & -A_{11}^{-1}A_{12}F_2 \\ -F_2A_{21}A_{11}^{-1} & F_2 \end{bmatrix}^{-1}$$

where $F_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$ and $F_2 = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$.

Applying this to our problem and focusing on the 2×2 matrix in the upper left of $(\tilde{Z}'\tilde{X})^{-1}$ (the other elements will have no impact on the bias due to the assumption that W and ϵ are uncorrelated), we get

$$(\tilde{Z}'\tilde{X})^{-1}\tilde{Z}'\epsilon = \begin{bmatrix} Z'X & Z'W \\ W'X & W'W \end{bmatrix}^{-1} \begin{bmatrix} Z'\epsilon \\ W'\epsilon \end{bmatrix}$$

$$(11) \qquad = \begin{bmatrix} [(Z'X) - Z'W(W'W)^{-1}W'Z]^{-1} & . \\ . & . \end{bmatrix} \begin{bmatrix} Z'\epsilon \\ W'\epsilon \end{bmatrix}$$

$$= \begin{bmatrix} (Z'M_W X)^{-1} & . \\ . & . \end{bmatrix} \begin{bmatrix} Z'\epsilon \\ W'\epsilon \end{bmatrix}$$

where $M_W \equiv I_R - W(W'W)^{-1}W'$ is the "residual maker" for W, i.e. it yields the residual from a projection of any variable onto W.[3]

Focusing on the first two elements of this matrix multiplication gives us the formulas for the bias on β_1 and β_2 that are analogous to those developed in the previous section:

$$(Z'M_W X)^{-1}Z'\epsilon = \begin{bmatrix} \frac{1}{N}X_1'M_W X_1 & \frac{1}{N}X_1'M_W X_2 \\ \frac{1}{N}Z'M_W X_1 & \frac{1}{N}Z'M_W X_2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N}X_1 M_W \epsilon \\ 0 \end{bmatrix}$$

Inverting the matrix manually generates a bias vector of:

$$(Z'M_W X)^{-1}Z'\epsilon = \begin{bmatrix} \frac{Z'M_W X_2}{(X_1'M_W X_1)(Z'M_W X_2) - (Z'M_W X_1)(X_1'M_W X_2)}(X_1 M_W \epsilon) \\ \frac{-Z'M_W X_1}{(X_1'M_W X_1)(Z'M_W X_2) - (Z'M_W X_1)(X_1'M_W X_2)}(X_1 M_W \epsilon) \end{bmatrix}$$

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

$$(12) \qquad bias = \frac{-Z'M_W X_1}{(X_1'M_W X_1)(Z'M_W X_2) - (Z'M_W X_1)(X_1'M_W X_2)}(X_1 M_W \epsilon)$$

This formula is analogous to what we had earlier (Equation (10)) once we partial out the influence of W. I'm sure it could also be derived by partialling W out of y, X_1, and X_2 and using the simple model if that is preferred.

4.4 Issues

1) Is the bound $SD(\epsilon_i) < SD(y_i)$ able to be motivated by any of the above arguments? I think we are ok in any case - I think the bounds can just be interpreted as a formal argument for how to pick the most "robust" instruments.

2) Extending to multiple x_{1i}'s

3) Extending to multiple instruments (best combination of instruments? - given the above result, it is not obvious that one would just want to use one instrument) - potentially with multiple

[3] **What about the overidentified case? Do we get an equally nice formula?**

endogenous variables (e.g. nested logit model).

4) One thing that I just thought of figured I should write it down - I was once convinced that our arguments sort of ruled out BLP/Bresnahan type instruments (because presumably the number of nearby competitors is not independent of one's own characteristics). This isn't necessarily the case - it depends on what sort of asymptotics one is thinking of. With a fixed number of markets (asymptotics in the number of products per market), this is likely a problem. But with the number of markets going to infinity (and a fixed number of products per market), I think these instruments could be fine.

5) Extending the bounds to non-linear models, e.g. of the form:

$$\delta(y_i, x_{1i}, x_{2i}; \beta_3) = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

like a BLP model. It is not completely clear to me how to do this, as one cannot use the "plim" arguments above. For proving consistency when z_i is uncorrelated with x_{1i} recall that we used the decomposition argument, i.e.

$$\begin{aligned}
\delta(Y, X_1, X_2; \beta_3) &= X_1\beta_1 + X_2\beta_2 + (P_{X_1} + I - P_{X_1})\epsilon \\
&= X_1(\beta_1 + \gamma) + X_2\beta_2 + (I - P_{X_1})\epsilon \\
&= X_1(\beta_1 + \gamma) + X_2\beta_2 + \epsilon^*
\end{aligned}$$

Then we set up a moment in ϵ^*, e.g.

$$E\left[Z'\epsilon^*\right] = 0$$

where Z contains both X_1 plus a set of other instruments Z_2 (to identify the parameters β_2 and β_3) i.e.,

$$E\begin{bmatrix} X_1'\epsilon^* \\ Z_2'\epsilon^* \end{bmatrix} = 0$$

By construction, X_1 is uncorrelated with ϵ^*. If Z_2 is uncorrelated with X_1 (as well as ϵ), then it must also be uncorrelated with ϵ^*. Hence, we have a consistent GMM estimator of $(\beta_1+\gamma, \beta_2, \beta_3)$.

But what if Z_2 is correlated with X_1? Restricting attention again to where both are scalars (e.g. the model $\delta(y_i, x_{2i}; \beta_2) = \beta_1 x_{1i} + \epsilon_i$), the expectation of the moment condition is not 0, but:

$$E\begin{bmatrix} X_1'\epsilon^* \\ Z_2'\epsilon^* \end{bmatrix} = E\begin{bmatrix} 0 \\ Z_2'(I - P_{X_1})\epsilon \end{bmatrix} = E\begin{bmatrix} 0 \\ Z_2'X_1(X_1'X_1)^{-1}X_1'\epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ Cov(z_{2i}, x_{1i})Cov(x_{1i}, \epsilon_i) \end{bmatrix}$$

I think the last equality holds???? Anyway, if so, by the same arguments as before, it seems like we can bound how "wrong" the moment condition is. It seems like one should be able to translate this into bounds on parameters. Suppose, e.g. $Cov(z_{2i}, x_{1i}) = 0.2$,and that we are willing to bound $Cov(x_{1i}, \epsilon_i)$ between -3 and 3. Then we know the second element of the moment condition is wrong by anywhere between -0.6 and 0.6. It seems like we could in theory just estimate the model for each quantity in this range, e.g. first use the moment condition:

$$E\left[\begin{array}{c} X_1'\epsilon^* \\ Z_2'\epsilon^* + 0.6 \end{array}\right] = 0$$

then, e.g.

$$E\left[\begin{array}{c} X_1'\epsilon^* \\ Z_2'\epsilon^* + 0.59 \end{array}\right] = 0$$

and so on until 0.6. This would presumably trace out bounds on the parameters, as long as there is some degree of smoothness relative to the step size one is using. If one could prove some sort of monotonicity result, one could probably just check 0.6 and -0.6, though I doubt this would be feasible in a non-linear model.

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.[4]

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 [Can generalize if wanted]. 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.

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

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.

[4] The data have a lot more, esp. the identity of offered networks for each bundle, demographic info in each market, etc. Keep it simple to start.

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 and Nevo 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: 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).[5]

Here are the preliminary results. First the results of the first-stage regression of price (tp) on all the explanatory variables and the instruments (Separate regressions for each instrument. Include instrumenting for price with itself for completeness).[6] Most of the results are of the correct sign and of reasonable magnitude. [Only weird ones is franchise fee - higher fees are associated with lower prices.]

Variable	fols	fivhp	fivfr~e	fivtcx	fivms~s	fivtip	fivti~t	fivti
tx	0.000	0.256	0.247	0.016	0.254	0.240	0.254	0.2
	0.000	0.009	0.009	0.021	0.009	0.009	0.010	0.0
tp	1.000							
	0.000							
hp		-0.051						

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

[6] 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.

		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								0.4
								0.0

Here are the associated IV results using each instrument as the single instrument for price (standard errors below the estimates). Also reported is the estimated impact to mean utility of an additional cable programming network.

As expected, instrumenting for price generally yields a larger estimated price sensitivity.

Variable	ols	ivhp	ivfra~e	ivtcx	ivmso~s	ivtip	ivtipst	ivtip
tp	-0.038	-0.022	-0.048	-0.024	-0.025	-0.070	-0.090	-0.0
	0.002	0.022	0.030	0.015	0.010	0.005	0.008	0.0
tx	0.029	0.025	0.032	0.026	0.026	0.038	0.042	0.0
	0.002	0.005	0.007	0.004	0.003	0.002	0.003	0.0

Here are the associated average estimated own-price elasticity (averaged across all products).[7] The patterns 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

[7] 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.

```
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
```

Here is an estimate of the relationship between each included instruments and the number of offered cable networks (tx). Note is statistically significant for all variables except the price of cable service at other systems within the same MSO (tip).

```
------------------------------------------------------------------------------------------------
    Variable |  xols       xhp     xfran~e    xtcx     xmsos~s    xtip     xtipst    xtipre
-------------+----------------------------------------------------------------------------------
          tp |   0.428
             |   0.017
          hp |             0.157
             |             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 |                                                                        -0.1
             |                                                                         0.0
------------------------------------------------------------------------------------------------
                                                                                  legend:
```

If we calculate an estimate of the upper bound on the absolute value of the bias using Equation (12) imposing the assumption $\frac{1}{N}X_1M_W\epsilon < \sqrt{\frac{1}{N}X_1M_WX_1}\sqrt{\frac{1}{N}yM_Wy}$ for each set of instruments, we get the following results.

```
. scalar list txtp vtx vy sdtx sdy biasnum2 ;
      txtp =  106516.91
       vtx =  448973.67
        vy =   7301.383
```

```
      sdtx =  670.05497
       sdy =   85.44813
  biasnum2 =  57254.945

. scalar list txtp tptp biasnumtp biasnum2 biasdenomtp biastp;
      txtp =  106516.91
      tptp =  249100.49
 biasnumtp =  106516.91
  biasnum2 =  57254.945
biasdenomtp =  1.005e+11
    biastp =  .06068658

. scalar list txhp tphp biasnumhp biasnum2 biasdenomhp biashp;
      txhp =  163200.41
      tphp = -10795.385
 biasnumhp =  163200.41
  biasnum2 =  57254.945
biasdenomhp = -2.223e+10
    biashp = -.42032582

. scalar list txfranfee tpfranfee biasnum2 biasnumfranfee biasdenomfranfee biasfranfee;
 txfranfee =  16825.949
 tpfranfee = -750.34934
  biasnum2 =  57254.945
biasnumfranfee =  16825.949
biasdenomfranfee = -2.129e+09
biasfranfee = -.45246954

. scalar list txtcx tptcx biasnumtcx biasnum2 biasdenomtcx biastcx;
     txtcx =   77317.13
     tptcx =  22386.574
biasnumtcx =   77317.13
  biasnum2 =  57254.945
biasdenomtcx =  1.815e+09
   biastcx =  2.4384633

. scalar list txmsosubs tpmsosubs biasnum2 biasnummsosubs biasdenommsosubs biasmsosubs;
 txmsosubs =  22107.897
```

```
 tpmsosubs = -28744.127
  biasnum2 =  57254.945
biasnummsosubs =  22107.897
biasdenommsosubs = -1.526e+10
biasmsosubs =  -.0829468

. scalar list txtip tptip biasnumtip biasnum2 biasdenomtip biastip;
     txtip = -2506.1517
     tptip =  67100.422
biasnumtip = -2506.1517
  biasnum2 =  57254.945
biasdenomtip =  3.039e+10
   biastip =  -.0047211

. scalar list txtipst tptipst biasnumtipst biasnum2 biasdenomtipst biastipst;
   txtipst = -17627.905
   tptipst =  35886.024
biasnumtipst = -17627.905
  biasnum2 =  57254.945
biasdenomtipst =  1.799e+10
 biastipst = -.05610395

. scalar list txtipreg tptipreg biasnumtipreg biasnum2 biasdenomtipreg biastipreg;
  txtipreg = -14663.474
  tptipreg =  35725.872
biasnumtipreg = -14663.474
  biasnum2 =  57254.945
biasdenomtipreg =  1.760e+10
biastipreg = -.04769696
```

Collecting just the overall bias terms, we get

```
. scalar list biastp biashp biasfranfee biastcx biasmsosubs biastip biastipst biastipreg ;
    biastp =  .06068658
    biashp = -.42032582
biasfranfee = -.45246954
   biastcx =  2.4384633
biasmsosubs =  -.0829468
   biastip =  -.0047211
```

```
 biastipst = -.05610395
biastipreg = -.04769696
```

6 Appendix

Consider the case when z_i^p has more than one element, e.g. $z_i^p = (z_{i1}^p, ..., z_{iK}^p)$. In this situation, we can do the above inversion sequentially, i.e.

$$
\begin{aligned}
z_{i1}^p &= F_1^{-1}(z_{i1}^{p2}|I_i^{t^d}) = h_1(z_{i1}^{p2}, I_i^{t^d}) \\
z_{i2}^p &= F_2^{-1}(z_{i2}^{p2}|I_i^{t^d}, z_{i1}^{p2}) = h_2(z_{i2}^{p2}, I_i^{t^d}, z_{i1}^{p2}) \\
&. \\
&. \\
z_{iK}^p &= F_2^{-1}(z_{iK}^{p2}|I_i^{t^d}, \{z_{ik}^{p2}\}_{k=1}^{k-1}) = h_K(z_{iK}^{p2}, I_i^{t^d}, \{z_{ik}^{p2}\}_{k=1}^{k-1})
\end{aligned}
$$

In this case, the variables $z_i^{p2} = (z_{i1}^{p2}, ..., z_{iK}^{p2})$ represent a set of components of z_i^p that are independent of $I_i^{t^d}$. With this decomposition, we can think of the firms information sets at the two points in time as:.

t^d - information set $I_i^{t^d} = \{m_i, \mu_i, z_i^d, z_i^{p1}\}$

t - information set $I_i^t = \{m_i, \mu_i, z_i^d, z_i^p\} = \{m_i, \mu_i, z_i^d, z_i^{p1}, z_i^{p2}\}$

where z_i^{p2} is independent of the other elements of the information set.

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