Orthogonal Instruments: Estimating Price Elasticities in the Presence of Endogenous Product Characteristics

Gregory S. Crawford

Dept. of Economics
University of Warwick

with Dan Ackerberg and Jin Hahn
UCLA

Université Toulouse I
November 24, 2009

Differentiated Product Demand Models

- Large area of research in empirical IO past 10-15 years has been models of differentiated product demand.
- Goal is to estimate a demand system for a differentiated product market
- This has many uses in industrial organization, marketing, strategy, e.g.
  - Own-price and cross-price elasticities, for pricing, merger analysis, etc.
  - Elasticities w.r.t. product characteristics
  - Welfare effects of new products or price or characteristic changes
  - Input into many other interesting IO questions

Differentiated Product Demand Models, cont.

- Typically have product level data on markets across time or space.
- Observe prices, characteristics, and market shares of products in each market.

A Dimensionality Problem

- Probably the biggest econometric hurdle in these models is a dimensionality problem:
  - With a homogenous product, there is one demand curve to estimate:

$$Q = \beta_0 + \beta_1 p + \varepsilon$$

  - With J differentiated products, there are J demand curves to estimate.

A Dimensionality Problem, cont.

- Even if one uses a linear demand system

$$\begin{aligned} Q_1 &= \beta_{0,1} + \beta_{1,1} p_1 + \cdots + \beta_{J,1} p_J + \varepsilon_1 \\ &\vdots \\ Q_J &= \beta_{0,J} + \beta_{1,J} p_1 + \cdots + \beta_{J,J} p_J + \varepsilon_J \end{aligned}$$

  unless J is very small there are typically too many parameters to estimate.

Solutions in the Literature

- Recent approaches reduce dimensionality by parameterizing elasticities based on observed product characteristics.
  1. Direct restrictions on coefficients in linear system
     - ★ Hausman (1996), Pinske, Slade, and Brett (2002), Davis (2006)
  2. Hedonic utility/aggregated discrete choice approach
     - ★ Bresnahan (1987), Berry, Levinsohn, and Pakes (1995) (BLP)
     - ★ Specify consumer utility functions as a function of a product's observed and unobserved characteristics.
     - ★ Aggregate demands over consumers to get product level market shares.
     - ★ Typically built up from individual level discrete choice models.

Solutions in the Literature, cont.

- Both approaches have advantages and disadvantages, although the second approach has arguably been more popular.
- We'll focus on the second approach, "Aggregated Discrete Choice Models",
  - But the basic ideas of our paper are also applicable to the first approach.

Types of Aggregated Discrete Choice Models

- Aggregated Discrete Choice Models are very common in the empirical literature.
- There are three main types:
  1. Logit Model
  2. Nested Logit Model
  3. Discrete-Choice Random Coefficients Model (RCM)
- RCM most flexible in terms of substitution patterns.

Logit Model

- Utility function (consumer i, product j)

$$u_{ij} = \beta p_j + x_j\theta + \xi_j + \varepsilon_{ij}$$

where

- x_j – observed (to econometrician) characteristics of product j
- p_j – price of product j
- ξ_j – unobserved characteristic of or demand shock for product j
- β, θ – parameters
- ε_{ij} – idiosyncratic taste consumer i has for product j (i.i.d Extreme Value)

Logit Model

- Consumer i chooses the product j that gives him/her the highest utility.
- Aggregating choices over consumers leads to the "market share" equation

$$\begin{aligned} s_j &= \int \cdots \int 1\left(u_{ij} > u_{ik} \quad \forall k \neq j\right) p\left(\varepsilon_1, \ldots, \varepsilon_J\right) \\ &= \frac{\exp\left[\beta p_j + x_j\theta + \xi_j\right]}{1 + \sum_k \exp\left[\beta p_k + w_k\theta + \xi_k\right]} \end{aligned}$$

Nested Logit

- Nested Logit Model (Goldberg (1995), Bresnahan, Stern, and Trajtenberg (1997))

$$u_{ij} = \beta p_j + x_j\theta + \zeta_{ig(j)} + \xi_j + \sigma\varepsilon_{ij}$$

where

  - $\zeta_{ig(j)}$ – consumer i's idiosyncratic taste for products in group g.

Discrete-Choice Random Coefficients

- Random Coefficients plus logit error (Berry, Levinsohn, and Pakes (1995))

$$u_{ij} = \beta_i p_j + x_j \theta_i + \xi_j + \varepsilon_{ij}$$

where

  - β_i – consumer i's distaste for price;
  - θ_i – consumer i's taste for characteristics.
  - Typically assume parameterized distributions for β_i and θ_i, e.g.

$$\beta_i \sim N\left(\beta, \sigma_\beta^2\right), \quad \theta_i \sim N\left(\theta, \Sigma_\theta\right)$$

Estimation

- Estimation of these models involves matching market shares predicted by the model to market shares observed in the data.
- This can often be quite straightforward, e.g.
  - Logit model generates an estimating equation of the form

$$\ln\left(\frac{s_j}{s_0}\right) = \beta p_j + x_j\theta + \xi_j$$

  - Nested Logit model

$$\ln\left(\frac{s_j}{s_0}\right) = \beta p_j + x_j\theta + \sigma\ln(s_{j|g}) + \xi_j$$

Estimation, cont.

- Random coefficients model is a bit more complicated
  - Estimating equation looks as follows:

$$\delta_j\left(\{s_l, w_l, p_l\}_{j=0}^{J}; \Sigma_\theta, \sigma_\beta^2\right) = \beta p_j + x_j\theta + \xi_j$$

  - Computing the left hand side variable typically requires simulation and an inversion routine.
- Estimation typically proceeds using either linear methods (logit, nested logit) or GMM (random coefficients models).

Estimation

- Researchers have typically worried about the possible endogeneity of price.
  - If the residual ξ_j represents unobserved product characteristics or unobserved demand shocks for product j,
  - Then a firm's profit maximizing price will generally depend on ξ_j,
  - Generating correlation and endogeneity.
- Estimation has typically proceeded using instruments for price,
  - Linear IV methods in logit and nested logit cases,
  - GMM with instruments for price in random coefficient models.

Price Instruments

- Commonly used instruments for price:
  - Cost shifters
  - Characteristics of competing products (BLP)
  - Prices of same product in other markets
    - Hausman (1996), Nevo (2001)

Exogenous Characteristics: Why?

- By contrast, researchers have (admittedly) relied upon the assumption that product characteristics x are exogenous.
- Question: Why is this?
  - Characteristics are choice variables just like price.
  - Seems like these choices might also depend on ξ_j.
- Answers:
  1. There is an argument is that price may be "more endogenous" than product characteristics
     - ★ As price is often is a more flexible and variable decision than are product characteristics (e.g. automobiles).
  2. Perhaps the problem is too hard to deal with?

Exogenous Characteristics: Problems

- We agree with the first argument, but
  1. This will clearly depend on the product under study.
  2. Even if x is "less endogenous" than p, it still may be problematic.
- Note also that if x is incorrectly assumed exogenous, it will generally bias *all* the coefficients in the model
  - Including the coefficient on price.
- This transmitted bias, e.g. to the price coefficient, might be expected to be less than any direct bias ...
  - (were one not to be instrumenting for price)
- But one can easily construct examples where the bias is large.

Exogenous Characteristics: Solutions in the Lit

A few solutions have been briefly discussed in the literature.

1. Find instruments for endogenous product characteristics.
   - Problems:
     - Already hard enough to find valid instruments for price
     - Unlike price, for which one often needs just one instrument, here one would need at least as many instruments as characteristics
     - If residual ξ_j is an unobserved product characteristic that is chosen by firms, it could be hard to find a valid instrument.

Exogenous Characteristics: Solutions in the Lit, cont.

A few solutions, cont:

2. BLP briefly suggest a solution based on timing. Suppose one has panel data, i.e. markets over time.

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \beta p_{jt} + w_{jt}\theta + \xi_{jt}$$

- Instead of considering a moment in ξ_{jt}, assume ξ_{jt} follows a first order Markov process and consider a moment in the innovations in ξ_{jt}, i.e.

$$E\left[\xi_{jt} - E\left[\xi_{jt} \middle| \xi_{jt-1}\right] \middle| w_{jt}, Z_{jt}\right]$$

Exogenous Characteristics: Solutions in the Lit, cont.

A few solutions, cont:

2. BLP Soln, cont.:
   - With appropriate assumptions on:
     1. The timing of the choice of product characteristics, and
     2. The information set of firms at various points in time

     One can show that this moment should equal zero.
   - Similar to Olley and Pakes (1996) methodology for estimating production functions.
     - ⋆ Reasonably demanding on the data, plus relies on fairly strong, non-directly-testable assumptions on unobservables.
     - ⋆ Applied in Sweeting (2007).

Exogenous Characteristics: Solutions in the Lit, cont.

A few solutions, cont:

3. Formally model endogenous choice of product characteristics
   - Recent paper by Crawford and Shum (2006)
   - Using results from screening literature
     - e.g. Mussa and Rosen (1978), Rochet and Stole (2002)
   - CS are able to explicitly model a multiproduct monopolist's choices of a one-dimensional product characteristic (and price).
   - Problems:
     - Very tied to assumptions of monopoly and that product characteristic space is one dimensional (or maybe discrete).
     - Would be much harder to do in oligopoly or with multidimensional characteristics. Lots of issues, including possible multiple equilibria.
     - Identification questions

Our "Solution"

- Certainly simpler that those described above
  - In some cases our "solution" will imply that existing estimation procedures provide consistent estimates of own- and cross-price elasticities,
  - Even if product characteristics are endogenous.
- Whether or not this is the case
  - i.e. whether existing procedures provide consistent estimates

  will actually be testable.
- If it is not the case, there may be alternatives

Our "Solution": Caveat

- One important caveat:
  - We will assume that our primary concern is estimation of own and cross price elasticities,
    - ★ i.e. we will give up on estimating elasticities w.r.t. characteristics.
    - ★ $\Rightarrow$ Only appropriate for answering price related (i.e. short run) policy questions.

Warm-up: OLS

- Our solution is based on a very simple econometric result...
- Consider a linear regression model

$$y_i = x_{i1}'\beta + x_{i2}'\theta + \varepsilon_i \tag{1}$$

  such that

  - x_{i1} is exogenous ($E[x_{i1}\varepsilon_i] = 0$), but
  - but x_{i2} is not ($E[x_{i2}\varepsilon_i] \neq 0$).

- Because the regressor vector $(x_{i1}', x_{i2}')'$ is not uncorrelated with the error ε_i,
  - A textbook argument establishes that the OLS estimator of the coefficient vector $(\beta', \theta')'$ is inconsistent in general.

When is OLS estimator for β consistent?

- We may ask if there are conditions under which the OLS estimator for β is consistent.
- For this purpose, write
$$\varepsilon_i = x_{i2}'\gamma + \varepsilon_i^*,$$
where
  - $\gamma = \left(E\left[x_{i2}x_{i2}'\right]\right)^{-1} E\left[x_{i2}\varepsilon_i\right]$ is the population regression coefficient when ε_i is regressed on x_{i2}.
- Note that $E\left[x_{i2}\varepsilon_i^*\right] = 0$ by construction.

When is OLS estimator for β consistent?, cont.

- We may then write

$$y_i = x_{i1}'\beta + x_{i2}'(\theta + \gamma) + \varepsilon_i^*. \tag{2}$$

- Question: If we regress y_i on $(x_{i1}', x_{i2}')'$,
  - Does the OLS estimator consistently estimate $\left(\beta', (\theta + \gamma)'\right)'$?
- Answer: Only if $E[x_{i1}\varepsilon_i^*] = 0$ and $E[x_{i2}\varepsilon_i^*] = 0$.
  - We are guaranteed that $E[x_{i2}\varepsilon_i^*] = 0$, but
  - $E[x_{i1}\varepsilon_i^*] = 0$ is likely to be violated in general

When is OLS estimator for β consistent?, cont.

- Why might $E[x_{i1}\varepsilon_i^*] \neq 0$
  - The new error term is $\varepsilon_i^* = \varepsilon_i - x_{i2}'\gamma$
  - As such, $E[x_{i1}\varepsilon_i^*] = E[x_{i1}\varepsilon_i] - E[x_{i1}x_{i2}']\gamma$
  - This will not be zero unless $E[x_{i1}x_{i2}'] = 0$ (which is testable)
- If $E[x_{i1}x_{i2}'] = 0$, then the OLS estimator consistently estimates $\left(\beta', (\theta+\gamma)'\right)'$.
  - In other words, the OLS estimator for β in the regression of y_i on $(x_{i1}', x_{i2}')'$ in (1) or (2) is consistent.
- **Bottom Line:** When x_{i1} and x_{i2} are uncorrelated, bias from an endogenous x_{i2} doesn't get transmitted to β

What of in IV Settings?

- It turns out that there is a similar result for IV models.
- We consider the similar linear regression model

$$y_i = x_{i1}'\beta + x_{i2}'\theta + \varepsilon_i, \tag{3}$$

where both x_{i1} and x_{i2} are endogenous.

  - In IO applications, $x_{i1} = p_i$, price, and $x_{i2} = x_i$, characteristics.

- Suppose
  - We have an instrument z_i for x_{i1} ($E[z_i\varepsilon_i] = 0$), but
  - No instrument for x_{i2}.

What of in IV Settings?

- We consider the properties of IV regression using z_i as instrument for x_{i1}, but incorrectly treating x_{i2} as exogenous.
- Because $(z_i', x_{i2}')'$ is not uncorrelated with ε_i,
  - We can easily see that the IV estimator for $(\beta', \theta')'$ is inconsistent in general.
- Our question is whether the estimator for β may be consistent under some conditions.
  - It turns out that if z_i is uncorrelated with x_{i2}, one will get consistent estimate of β, even with endogenous x_{i2}.

What of in IV Settings?

- In order to understand this result, we again write

$$\varepsilon_i = x_{i2}'\gamma + \varepsilon_i^*,$$

where ε_i^* denotes the residual in the projection ε_i on x_{i2}.

- Now, we rewrite the model

$$y_i = x_{i1}'\beta + x_{i2}'(\theta + \gamma) + \varepsilon_i^* \tag{4}$$

Note

- x_{i2} is uncorrelated with ε_i^* $(E[x_{i2}\varepsilon_i^*] = 0)$ by construction.

What of in IV Settings?

- Note also that

$$E\left[z_i\varepsilon_i^*\right] = E\left[z_i\left(\varepsilon_i - x_{i2}'\gamma\right)\right] = E\left[z_i\varepsilon_i\right] - E\left[z_i x_{i2}'\right]\gamma = 0$$

  if $E\left[z_i x_{i2}'\right] = 0$.

- Following the identical logic as in the OLS case,
- It follows that the IV regression of y_i on $(x_{i1}', x_{i2}')'$ using $(z_i', x_{i2}')'$
  - Will produce a consistent estimator of $\left(\beta', (\theta+\gamma)'\right)'$.

A (Small) Efficiency Result

- Note that we one can rewrite the model as

$$y_i = x_{i1}'\beta + u_i,$$

where $u_i \equiv x_{i2}'\theta + \varepsilon_i$.

  - If $E[z_i x_{i2}'] = 0$,
  - Then $u_i \equiv x_{i2}'\theta + \varepsilon_i$ satisfies $E[z_i u_i] = E[z_i x_{i2}']\theta + E[z_i \varepsilon_i] = 0$,

- This suggests that β could be consistently estimated by fitting this "smaller" equation with z_i as an instrument.
- It turns out the asymptotic variance of this smaller IV regression is bigger than the one in the larger IV regression.
  - So you are better off including x_{i2} in the model.

A Useful Result?

- This seems to us to be much more useful than the OLS result
  - With the OLS result, either x_{i1} and x_{i2} are correlated or they are uncorrelated.
    - There is not much one can do if they are correlated.
  - With the IV result, one often has a choice of what instrument z_i to use.
    - Appropriate choice of instruments may lead to a desirable result.
- In our demand system context, we are suggesting looking for price instruments, z that are uncorrelated with the potentially endogenous product characteristics, x.
  - A nice aspect of this condition is that it is testable, since both z and x are observed.

Applications to DCMs in IO: Estimation

- Let's formally apply this result to our aggregated discrete choice models.
- Start with the Logit model:

$$\ln\left(\frac{s_j}{s_0}\right) = \beta p_j + x_j\theta + \xi_j$$

- Assume both p_j and x_j are endogenous, but we only have an instrument z_j for p_j.
  - Run IV using z_j as instrument for p_j but treating X_j as exogenous.
- Our previous result says that this will generate consistent estimates of the price coefficient β (but not the characteristics coefficients θ) if z_j is uncorrelated with x_j.

Applications to DCMs in IO: Elasticities

- But... we are not done yet.
  - We are typically not interested in the price coefficient in utility, β, per-se...
  - We are interested in own- and cross-price elasticities and derivatives.
- Fortunately, in the logit model, one can easily derive

$$\frac{\partial s_j}{\partial p_j} = -\beta s_j(1 - s_j)$$

$$\frac{\partial s_j}{\partial p_k} = \beta s_j s_k$$

- Thus, all own and cross price elasticities and derivatives can be written as just functions of the data (s and p) and β.
  - In other words, we do not need to consistently estimate θ to obtain these elasticities.

Elasticities in the Nested Logit

What about the Nested Logit model?

$$\ln\left(\frac{s_j}{s_0}\right) = \beta p_j + x_j\theta + \sigma \ln(s_{j|g}) + \xi_j$$

- In the nested logit model, there is another endogenous right hand side variable, $\ln(s_{j|g})$.
  - Hence one always needs an additional instrument as compared to the logit model.
- Various possibilities here
  - One natural additional instrument would be the price shifters of other products in your "group", which should be correlated with $\ln(s_{j|g})$.
- In any case, denote the instruments as z_{1j} and z_{2j}.
- Again, our earlier results implies that as long as both z_{1j} and z_{2j} are uncorrelated with the product characteristics x_j, one will obtain consistent estimates of β and σ.

Elasticities in the Nested Logit

- Again it also turns out that own and cross price elasticities can be written as only functions of the data (s and p), β, and σ, i.e.

$$\frac{\partial s_j}{\partial p_j} = -\beta s_j \left(\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} s_{j|g} - s_j \right)$$

$$\frac{\partial s_j}{\partial p_k} = \beta s_k \left(\frac{\sigma}{1-\sigma} s_{j|g} + s_j \right) \qquad \text{if } j \text{ and } k \text{ in same group}$$

$$\frac{\partial s_j}{\partial p_k} = \beta s_k s_j \qquad \text{if } j \text{ and } k \text{ in different groups}$$

- So we can again consistently estimate these quantities if our price instruments are uncorrelated with the endogenous product characteristics.

Elasticities in the RCM

The Random Coefficient model is a bit more complicated:

$$\delta_j \left(\{s_l, w_l, p_l\}_{j=0}^{J} ; \Sigma_\theta, \sigma_\beta^2 \right) = \beta p_j + x_j \theta + \xi_j$$

- Estimation of these models typically proceeds using GMM, starting from the following orthogonality condition.

$$E \left[\xi_j \otimes \begin{pmatrix} z_j \\ x_j \end{pmatrix} \right] = 0$$

- This implies that

$$E \left[\left(\delta_j \left(\{s_l, x_l, p_l\}_{l=0}^{J} ; \Sigma_\theta, \sigma_\beta \right) - \beta p_j - x_j \theta \right) \otimes \begin{pmatrix} z_j \\ x_j \end{pmatrix} \right] = 0$$

at the true parameters.

Elasticities in the RCM

- Again note that more instruments are required here (to identify σ and Σ)
- If x_j is endogenous, this moment condition doesn't hold, but, as above, we can linearly project ξ_j on x_j to get:

$$\delta_j \left(\{s_l, X_l, p_l\}_{l=0}^{J}; \Sigma_\theta, \sigma_\beta \right) = \beta p_j + x_j (\theta + \gamma) + \xi_j^*$$

- By construction, ξ_j^* is uncorrelated with x_j.
- The question is whether ξ_j^* is uncorrelated with the instruments z_j.
- As above, given that ξ_j^* is uncorrelated with the instruments z_j, this will be the case if the instruments z_j are uncorrelated with the product characteristics x_j.
- Hence, estimation can proceed using the moment

$$E\left[\xi_j^* \otimes \begin{pmatrix} z_j \\ x_j \end{pmatrix}\right] = 0$$

Elasticities in the RCM

- Again, we need to show that what we actually want, i.e. price derivatives and elasticities, do not depend on an estimate of θ.
- This is a bit tougher, but can be shown using the "inversion" of Berry (1994) and Berry, Levinsohn, and Pakes (1995)
- Goal is to show that we can write price derivatives/elasticities as:

$$\frac{\partial s_j}{\partial p_k} = f_{j,k}\left(data, \Sigma_\theta, \sigma_\beta, \beta\right)$$

  i.e. can be written as a function of the data and parameters that can be consistently estimated (i.e. can be computed without knowing θ)
- The Berry/BLP inversion shows that in this class of models:

$$\beta p_j + x_j\theta + \xi_j \equiv \delta_j = h\left(data, \Sigma_\theta, \sigma_\beta\right)$$

Elasticities in the RCM

- Hence, it suffices to show that we can write price derivatives as:

$$\frac{\partial s_j}{\partial p_k} = f_{j,k}\left(\{\delta_l\}_{l=0}^{J}, data, \Sigma_\theta, \sigma_\beta, \beta\right)$$

- Now, consider the RCM utility function:

$$\begin{aligned} u_{ij} &= \beta_i p_j + x_j\theta_i + \xi_j + \varepsilon_{ij} \\ &= \left(\beta + \tilde{\beta}_i\right) p_j + x_j\left(\theta + \tilde{\theta}_i\right) + \xi_j + \varepsilon_{ij} \\ &= \beta p_j + x_j\theta + \xi_j + \tilde{\beta}_i p_j + x_j\tilde{\theta}_i + \varepsilon_{ij} \\ &= \delta_j + \tilde{\beta}_i p_j + x_j\tilde{\theta}_i + \varepsilon_{ij} \end{aligned}$$

- Hence, θ only enters the model through δ, and both market shares and market share derivatives w.r.t. p do not depend on θ (given δ's)

Elasticities in the RCM

- Summary of argument:
  - θ's only enters the price elasticities/derivatives through δ's.
  - δ's can be written as a function of the data given parameters $(\beta, \sigma_\beta, \Sigma_\theta)$.
  - Thus, given we can consistently estimate $(\beta, \sigma_\beta, \Sigma_\theta)$, we can compute price elast/derivs without knowing θ.

Plan for the rest of the talk

- We've now established our basic results about the merits of Orthogonal Instruments
  - Particularly as applied to ADMs commonly used in empirical IO
- The balance of the talk
  - Introduces some of the various theory extensions we've obtained
  - Presents some preliminary results applying these ideas in US pay-television markets

A Counter-Intuitive Efficiency Result

- We can show a rather counter-intuitive result on orthogonal instruments:
  - Even if we have valid instruments on the endogenous regressor x_{i2}, it may make more sense to proceed as if x_{i2} were exogenous because we may gain efficiency as a result.
- Using (z_{i1}, z_{i2}) as instruments, we can show the asymptotic variance for the estimator for β is

$$E\left[\varepsilon_i^2\right]\left(\pi_{11}' E\left[z_{i1}z_{i1}'\right]\pi_{11}\right)^{-1} \tag{5}$$

- Using (z_{i1}, x_{i2}) as instruments, we can similarly show it is

$$E\left[\left(\varepsilon_i^*\right)^2\right]\left(\pi_{11}' E\left[z_{i1}z_{i1}'\right]\pi_{11}\right)^{-1}. \tag{6}$$

  where $\varepsilon_i = x_{i2}'\gamma + \varepsilon_i^*$

- Because $E\left[\varepsilon_i^2\right] \geq E\left[\left(\varepsilon_i^*\right)^2\right]$ by construction, we conclude that it makes sense to adopt the operating assumption that x_{i2} is exogenous.

Non-Parametric Extension

- The preceding discussion can be generalized to nonparametric context immediately.
- For this purpose, consider the following non-parametric model

$$y = g\left(x_1, x_2, \varepsilon\right)$$

  where x_1 is endogenous.
    - We allow x_2 to be endogenous, but
    - Consider identification under the assumption that the instrument z is independent of (x_2, ε).
- Repeating the similar logic as in the previous sections,
    - Let F denote the conditional quantile of ε given (x_2, z), and
    - Let $\varepsilon^* = F\left(\varepsilon \mid x_2, z\right)$.

Non-Parametric Extension

- Because z is independent of (x_2, ε), we can write without loss of generality
  - $F(\varepsilon | x_2, z) = F(\varepsilon | x_2)$, or $\varepsilon^* = F(\varepsilon | x_2)$.
- By construction, ε^* is independent of (x_2, z).
- Note that $\varepsilon = F^{-1}(\varepsilon^* | x_2)$, $\Rightarrow$

$$y = g(x_1, x_2, \varepsilon) = g\left(x_1, x_2, F^{-1}(\varepsilon^* | x_2)\right) \equiv g^*(x_1, x_2, \varepsilon^*)$$

such that ε^* is independent of (x_2, z).

Non-Parametric Extension

- If we assume that g is monotonic in ε, then monotonicity of F implies that g^* is monotonic in ε^* as well.
- With the additional assumption that g^* is *strictly* monotonic in ε^*, we can use the result by Chernozukov, Imbens, and Newey (2007) to conclude:
$$\frac{\partial g^* \left(x_1, x_2, \varepsilon^*\right)}{\partial x_1} = \frac{\partial g \left(x_1, x_2, \varepsilon\right)}{\partial x_1}$$
is identified.
- Because the $F \left(\varepsilon | x_2 \right)$ is not identified, the only thing we identify is the causal effect of x_1 at any quantile of ε^* and any given value of x_2.
  - This has some significance in IO applications because it means that we can estimate the causal effect of x_1 on y for any observation in the "dataset", since we have an estimate of each observation's ε^*.

What Types of Instruments Might be Orthogonal?

- In our demand system context, is there a reason to think one might be able to find price instruments that are uncorrelated with product characteristics?
- We hope so...
  - Recall that one interpretation of ξ is that it represents product characteristics that are observed by firms and customers but unobserved to the econometrician.
    - Standard IV condition is that z is uncorrelated with these unobserved product characteristics.
  - If we can find z's that are uncorrelated with these unobserved product characteristics...
    - Shouldn't we be able to find z's that are uncorrelated with the observed product characteristics?

What Types of Instruments Might be Orthogonal?, cont.

What sort of data generating processes would generate such instruments?

- We are still thinking about these issues, but can describe one particular process.
  - Want instruments that affect price-setting but do not affect choices of characteristics.
  - Perhaps the easiest way to think of such an instrument is to think of a timing story.
    - Suppose product characteristics are chosen at some point in time prior to when price is set.
    - Then what we optimally would want would be shocks that occur between these points in time and that are unanticipated by firms.
    - For example, unanticipated shocks to input prices that occur between these points in time would be excellent instruments. Could use commodity or exchange rate futures markets.

What Types of Instruments Might be Orthogonal?, cont.

DGPs to generate orthogonal instruments, cont:

- This timing story is somewhat reminiscent of the Olley and Pakes (1996)-style identification strategy, but in contrast to that, this is a directly *testable* restriction.
- Currently thinking through what the implications are on various types of instruments, e.g. standard cost shifters, BLP "competitive" instruments and Hausman/Nevo "other price" instruments.
  - Likely depends on the interpretation of ξ.

Uses Beyond Industrial Organization

- Seems to us that Orthogonal Instruments may also be useful for general IV situations
- Seems quite common to be interested in a subset of the structural parameters, e.g.
  - Returns to education but not to experience, tenure, etc.
- Examining correlations between instruments and "exogenous" variables can tell you how robust your estimates are to those "exogenous" variables actually being endogenous.
- Also may provide a way of choosing between instruments.

Bounding the Bias

- What can we say when we cannot find any instrument that is orthogonal to a potentially-endogenous x_2?
  - We can't use our consistency and efficiency results
  - We can, however, try to bound the magnitude of any bias
- Suppose, for the model

$$y = x_1\beta + x_2\theta + \varepsilon$$

  - We have instruments z_1 and z_2 on x_1.
  - We are concerned that x_2 may be endogenous as well, but we don't have an instrument for x_2.

Bounding the Bias, cont.

- We compare the asymptotic bias for β of the two IV estimators, one using z_1 and the other one using z_2.
- Let $\hat{\beta}_j$, $j = \{1, 2\}$ be the estimator of β using z_j as an instrument
- The asymptotic bias for each estimator is then

$$\text{plim}\,\widehat{\beta}_1 - \beta = \frac{-E\left[z_1 x_2\right]}{E\left[z_1 x_1\right] E\left[x_2^2\right] - E\left[x_2 x_1\right] E\left[z_1 x_2\right]} E\left[x_2 \varepsilon\right]$$

$$\text{plim}\,\widehat{\beta}_2 - \beta = \frac{-E\left[z_2 x_2\right]}{E\left[z_2 x_1\right] E\left[x_2^2\right] - E\left[x_2 x_1\right] E\left[z_2 x_2\right]} E\left[x_2 \varepsilon\right]$$

  - If, indeed, x_2 is uncorrelated with ϵ, then there is no bias.
  - Else, as usual, that bias is transmitted to β

Bounding the Bias: Intermediate Results

- It turns out we can simplify this.
- We can use the identity

$$\frac{-E\left[z_1 x_2\right]}{E\left[z_1 x_1\right] E\left[x_2^2\right]-E\left[x_2 x_1\right] E\left[z_1 x_2\right]}=-\frac{\frac{E\left[z_1 x_2\right]}{E\left[x_2^2\right]}}{E\left[z_1 x_1\right]-\frac{E\left[x_2 x_1\right]}{E\left[x_2^2\right]} E\left[x_2^2\right] \frac{E\left[z_1 x_2\right]}{E\left[x_2^2\right]}}=-\frac{\lambda_1}{E\left[z_1 x_1\right]-E\left[\left(\mu x_2\right)\left(\lambda_1 x_2\right)\right]}$$

where

$$\lambda_1=\frac{E\left[x_2 z_1\right]}{E\left[x_2^2\right]}, \quad \mu=\frac{E\left[x_2 x_1\right]}{E\left[x_2^2\right]}$$
$$v_1=z_1-\lambda_1 x_2, \quad u=x_1-\mu x_2$$

Note:

- $\lambda_j =$ the correlation between our x_1-instrument and x_2 (ideally this would be zero)
- u, $v_j = x_1$, z_j, controlling for x_2.

Bounding the Bias: Intermediate Results, cont.

- Noting that

$$E\left[z_1 x_1\right] - E\left[(\mu x_2)(\lambda_1 x_2)\right] = E\left[(\lambda_1 x_2 + v_1)(\mu x_2 + u)\right] - E\left[(\mu x_2)(\lambda_1 x_2)\right] = E\left[u v_1\right]$$

we can conclude that the ratio in the bias formula is just

$$\frac{-E\left[z_1 x_2\right]}{E\left[z_1 x_1\right] E\left[x_2^2\right] - E\left[x_2 x_1\right] E\left[z_1 x_2\right]} = -\frac{\lambda_1}{E\left[u v_1\right]}$$
$$\frac{-E\left[z_2 x_2\right]}{E\left[z_2 x_1\right] E\left[x_2^2\right] - E\left[x_2 x_1\right] E\left[z_2 x_2\right]} = -\frac{\lambda_2}{E\left[u v_2\right]}$$

Bounding the Bias: An Intuitive Formula

$$\text{plim}\, \widehat{\beta}_1 - \beta = \frac{-E\left[z_1 x_2\right]}{E\left[z_1 x_1\right] E\left[x_2^2\right] - E\left[x_2 x_1\right] E\left[z_1 x_2\right]} E[x_2 \varepsilon] = -\frac{\lambda_1}{E\left[u v_1\right]} E[x_2 \varepsilon]$$

- In other words, the asymptotic bias depends on three factors
  - The correlation between our instrument and the potentially endogenous variable (λ_j)
  - The strength of our instrument for our variable of interest, controlling for x_2 ($E[uv_j]$), and
  - The correlation between our the potentially endogenous variable and the error, $E[x_2\varepsilon]$

Bounding the Bias: How to use?

We can use our bias results in a number of ways:

1. Selecting an instrument to minimize asymptotic bias:
    - In particular, we may want to consider choosing an instrument depending on whether

$$\left|\frac{\lambda_1}{E\left[uv_1\right]}\right| \lesseqgtr \left|\frac{\lambda_2}{E\left[uv_2\right]}\right|$$

Bounding the Bias: How to use?

How to use our bias results, cont.:

2. Minimizing the asymptotic bias using a linear combination of instruments
   - Let $\delta_1 z_1 + \delta_2 z_2$ be an instrument.
   - We then have the asymptotic bias proportional to
   $$-\frac{\delta_1 \lambda_1 + \delta_2 \lambda_2}{\delta_1 E\left[uv_1\right] + \delta_2 E\left[uv_2\right]}$$
   - We can eliminate the asymptotic bias if $\delta_1 = 1$ and $\delta_2 = -\frac{\lambda_1}{\lambda_2}$.
   - Unfortunately, our efficiency results are for orthogonal instruments, not for "estimated orthogonal instruments".
     - ★ If we have two instruments, might we not instead just do "vanilla IV", i.e. instrument for x_1 and x_2 with z_1 and z_2?

Bounding the Bias: How to use?

How to use our bias results, cont.:

3. Perhaps we can bound the absolute magnitude of the bias?
   - If we are willing to make an assumption on ε,
     - For example, that $sd(\varepsilon) < sd(y)$
     - (As would be true as long as the explanatory variables and ε are not too negatively correlated)
   - Then

$$\begin{aligned} abs(cov(x_2, \varepsilon)) &< sd(x_2)sd(\varepsilon) \\ &< sd(x_2)sd(y) \end{aligned}$$

   - And we can bound the bias:

$$abs(bias) < \frac{\lambda_1}{E\left[uv_1\right]} sd(x_2)sd(y)$$

Empirical Application: US Pay-TV Markets

- Data on demand for cable systems. Goal is to estimate price elasticity of demand.
  - e.g. To measure cable system market power;
  - How market power has changed in response to satellite competition, etc.
- Observe prices, service characteristics, and market shares for cross section of approximately 4000 cable systems across the US.

Empirical Application: US Pay-TV Markets

- Keep things simple. We consider a logit demand model:

$$u_{ij} = X_j\beta - \alpha p_j + W_j\gamma + \xi_j + \varepsilon_{ij}$$

  - We only consider one service characteristic X_j – the number of cable programming networks offered
  - W_j are other explanatory variables that are assumed exogenous.
- In a given market, cable system may offer a number of alternative products j (e.g. basic, expanded basic) characterized by different prices and number of networks.

Empirical Application: US Pay-TV Markets

- We consider a number of potential instruments for price:
  1. hp – Homes Passed – the number of homes potentially served by the system. May create bargaining power with television networks.
  2. franfee – Franchise Fees – fees paid to local governing bodies in return for access to streets to deliver service.
  3. tcx – Average Affiliate Fees – average fees charged by networks on a particular cable system.
  4. msosubs – Multiple System Operator (MSO) Subscribers - many operators own multiple cable systems across the country (e.g. Comcast, Cox).
     - This is the total number of subscribers on an operator's systems. Again, this could affect bargaining power.
  5. tip, tipst, tipreg – prices in other markets (ala Hausman (1996) and Nevo (2001)) of the same MSO.
     - Idea is that this will pick up supply shocks.

First-Stage Results

- First stage results (all instruments used separately):

Instrument	Coefficient
hp	-0.51 (0.06)
franfee	-2.73 (0.047)
tcx	1.287 (0.110)
msosubs	-0.337 (0.019)
tip	0.642 (0.017)
tipst	0.487 (0.02)
tipreg	0.454 (0.02)

  - All highly significant (though no clustering)
  - All except for franfee are the anticipated sign.

Correlation of Instruments with Product Characteristic

- Regression of product characteristic on the various instruments plus bias bounds.

	Regression coef	Bound on abs bias
Hp	0.157 (0.008)	0.42
Franfee	0.935 (0.064)	0.45
Tcx	4.698 (0.030)	2.43
Msosubs	0.217 (0.027)	0.08
Tip	-0.024 (0.027)	0.004
Tipst	-0.213 (0.030)	0.06
Tipreg	-0.169 (0.029)	0.05

  - Suggests that tip may be the best instrument – insignificant regression coefficient and very small bias.

Estimated Price Coefficients

- Estimated price coefficients using each of the instruments separately

	Price Coefficient
OLS	-0.038 (0.002)
Hp	-0.022 (0.022)
Franfee	-0.048 (0.030)
Tcx	-0.024 (0.015)
msosubs	-0.025 (0.010)
Tip	-0.070 (0.005)
Tipst	-0.090 (0.008)
Tipreg	-0.078 (0.008)

  - Fairly large differences across specifications. Implied elasticities between -0.24 and -1. General consensus is that elasticities are closer to -1.
  - Tip related instruments provide the most reasonable estimates,

Conclusions

- Perhaps endogenous product characteristics in differentiated product demand models is not as problematic as commonly thought.
- We derive conditions under which we can show that standard estimation procedures provide consistent estimates of price derivatives and elasticities
  - These conditions are testable and have implications on what price instruments one might want to be using use in practice.
- Also sheds light on what sort of data-generating processes would be most likely to generate such instruments.
- Idea seems to work reasonably well in a simple example.

Next Steps

- Extend the data and analysis to more systems, years, etc.
  - As in Crawford and Yurukoglu (2009)
- Further develop our thinking about the timing of decisions in pay-television markets
  - Seems reasonable that number of channels “more exogenous” in the short-run than prices
  - In which case can use *changes* in costs / other prices / similar to identify likely-to-be-orthognoal instruments.

BERRY, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics*, 25(2), 242–262.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

BRESNAHAN, T. (1987): "Competition and Collusion in the American Auto Industry: The 1955 Price War," *Journal of Industrial Economics*, 35(4), 457–482.

BRESNAHAN, T., S. STERN, AND M. TRAJTENBERG (1997): "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the late 1980s," *Rand Journal of Economics*, pp. S17–S44.

CHERNOZUKOV, V., G. IMBENS, AND W. NEWEY (2007): "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139(1), 4–14.

CRAWFORD, G., AND M. SHUM (2006): "Empirical Modeling of Endogenous Quality Choice: The Case of Cable Television," mimeo, University of Arizona.

Crawford, G., and A. Yurukoglu (2009): "The Welfare Effects of Bundling in Multichannel Television," mimeo, University of Warwick.

Davis, P. (2006): "Spatial Competition in Retail Markets: Movie Theaters," *Rand Journal of Economics*, 37(4), 964–982.

Goldberg, P. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry," *Econometrica*, 63, 891–951.

Hausman, J. (1996): "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. Bresnahan, and R. Gordon. University of Chicago Press.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–17.

Nevo, A. (2001): "Measuring Market Power in the Ready-To-Eat Cereal Industry," *Econometrica*, 69(2), 307–342.

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

Pinske, J., M. Slade, and C. Brett (2002): "Spatial Price Competition," *Econometrica*, 70(3), 1111–1153.

Rochet, J.-C., and L. Stole (2002): "Nonlinear Pricing with Random Participation," *Review of Economic Studies*, 69(1), 277–311.

Sweeting, A. (2007): "Dynamic Product Repositioning in Differentiated Products Industries: The Case of Format Switching in the Commercial Radio Industry," Working Paper, Duke University.