Week 3: Demand Estimation

Chris Conlon

Grad IO

October 2, 2018

Heterogeneity and Endogeneity

Putting it Together

▶ Now we want to have both price endogeneity and flexible substitution in the same model.

▶ We are ultimately going with the random coefficients logit model, but we will start with the logit and nested logit.

Basic Idea from Price Endogeneity

$$s_{jt} = \int \frac{\exp[x_{jt}\beta_i]}{1 + \sum_k \exp[x_{kt}\beta_i]} f(\beta_i|\theta)$$

- We know prices are set with demand in mind and this can create an endogeneity problem.
- How do we deal with it?
- We would like to instrument in this world but what is the error term exactly?
- An obvious choice might be $\eta_{jt} = (s_{jt}(\theta) - \tilde{s}_{jt})$
- Can we find things that are orthogonal to the error between observed and predicted market shares?
- Do we have the usual IV conditions (exogeneity, relevance, monotonicity, etc.)

Basic Idea from Price Endogeneity

▶ We need to add an unobservable quality term ξ_{jt} to our model

$$
\begin{aligned}
u_{ijt} &= x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ij} \\
s_{jt} &= \int \frac{\exp[x_{jt}\beta_i + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta_i + \xi_{kt}]} f(\beta_i|\theta)
\end{aligned}
$$

▶ The idea is that ξ_{jt} is observed to the firm when prices are set, but not to us the econometricians.

▶ We call ξ_{jt} a vertical component, because all consumers agree on its value.

▶ This allows for products j to better than some other product in a way that is not fully explained by differences in x_j and x_k.

▶ Basically there is something about a BMW that makes it better than a Peugeot in a way that is not fully captured by its mileage, weight, horsepower, etc. that leads to it having higher sales and/or higher prices.

Inversion: IIA Logit

▶ Think about the plain IIA logit for a minute:

$$
\begin{aligned}
u_{ijt} &= x_{jt}\beta + \xi_{jt} + \varepsilon_{ij} \\
s_{jt} &= \frac{\exp[x_{jt}\beta + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]}
\end{aligned}
$$

▶ Take logs

$$
\begin{aligned}
\ln s_{0t} &= -\log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right) \\
\ln s_{jt} &= [x_{jt}\beta + \xi_{jt}] - \log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right) \\
\ln s_{jt} - \ln s_{0t} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt}
\end{aligned}
$$

Inversion: IIA Logit

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{data!} \;=\; \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}}$$

- ▶ The LHS is data! The RHS is now a linear IV problem!
- ▶ α the price coefficient is the endogenous parameter.
- ▶ We know how to solve this. We need instruments that shift p_{jt} but are orthogonal to ξ_{jt}.
- ▶ Economic theory tells us how: cost shifters, markup shifters.
- ▶ Markups in IIA logit are pretty boring since they only depend on your shares and α.
- ▶ If number of products varies across markets, that works. Otherwise you want cost shifters in cross section or time series.

Was that magic?

- No. It was just a nonlinear change of variables from $\eta_{jt} \to \xi_{jt}$.
- Our moment condition is just that $E[\xi_{jt}|x_{jt}, z_{jt}] = 0$.
- We moved from the space of shares and MLE for the logit to the space of utilities and an IV model.
- We are losing some efficiency – but now we are able to estimate under weaker conditions.

Caveats

▶ We do need a technical condition. This only works if the market size $N \to \infty$.

▶ That is our data/shares we must believe we are observing without any sampling error.

▶ This is not necessary for the multinomial MLE where shares have some natural sampling variation.

▶ In our IV/GMM approach we cannot have this sampling error. (Why?).

Inversion: Nested Logit (Berry 1994 / Cardell 1991)

This takes a bit more algebra but not much

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{data!} = x_{jt}\beta - \alpha p_{jt} - \sigma \underbrace{\log(s_{j|gt})}_{data!} + \xi_{jt}$$

- Same as logit plus an extra term $\log(s_{j|g})$ the within group share.
- We now have a second endogenous parameter.
- If you don't see it – realize we are regressing Y on a function of Y. This should always make you nervous.
- If you forget to instrument for σ you will get $\sigma \to 1$ because of attenuation bias.
- A good instrument for σ is the number of products within the nest. Why?

Inversion: BLP

We can't solve for δ_{jt} directly this time. We often exploit a trick when β_i, ν_i is normally distributed:

$$s_{jt} = \int \frac{\exp[\delta_{jt} + x_{jt} \cdot \Sigma \cdot \nu_i]}{1 + \sum_k \exp[\delta_{kt} + x_{kt} \cdot \Sigma \cdot \nu_i]} f(\nu_i | \theta)$$

- This is a $J \times J$ system of equations for each t.
- It is diagonally dominant.
- There is a unique vector ξ_t that solves it for each market t.
- If you can work out $\frac{\partial s_{jt}}{\partial \delta_{kt}}$ (easy) you can solve this using Newton's Method.

Contraction: BLP

BLP actually propose an easy solution to find δ_t. Fix θ and solve for δ. Think about doing this one market at a time:

$$\delta^{(k)}(\theta) = \delta^{(k-1)}(\theta) + \log(\tilde{s}_j) - \log(s_j(\delta_t^{(k-1)}, \theta))$$

▶ They prove (not easy) that this is a contraction mapping.

▶ If you keep iterating this equation enough
$|\delta^{(k)}(\theta) - \delta^{(k-1)}(\theta)| < \epsilon_{tol}$ you can recover the δ's so that the observed shares and the predicted shares are identical.

▶ Practical tip: ϵ_{tol} needs to be as small as possible. ($\approx 10^{-13}$).

▶ Practical tip: Contraction isn't as easy as it looks:
$\log(s_j(\delta_t^{(k-1)}, \theta))$ requires computing the numerical integral each time (either via quadrature or monte carlo).

BLP Pseudocode

From the outside, in:

- ▶ Outer loop: search over nonlinear parameters θ to minimize GMM objective:

$$\widehat{\theta_{BLP}} = \arg \max_\theta (Z'\hat{\xi}(\theta))W(Z'\hat{\xi}(\theta))'$$

- ▶ Inner Loop:
    - ▶ Fix θ.
    - ▶ Solve for δ so that $s_{jt}(\delta, \theta) = \tilde{s}_{jt}$.
        - ▶ Computing $s_{jt}(\delta, \theta)$ requires numerical integration (quadrature or monte carlo).
    - ▶ We can do IV-GMM to recover $\hat{\alpha}(\theta), \hat{\beta}(\theta), \hat{\xi}(\theta)$.

    $$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

    - ▶ Use $\hat{\xi}(\theta)$ to construct moment conditions.
- ▶ When we have found $\hat{\theta}_{BLP}$ we can use this to update $W \to W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

BLP Estimation

▶ Now that you have done change of variables to get:

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

▶ We can do IV-GMM to recover $\hat{\alpha}(\theta), \hat{\beta}(\theta), \hat{\xi}(\theta)$.

▶ Outer Loop update guess θ, solve for δ and repeat.

$$\widehat{\theta_{BLP}} = \arg\max_{\theta}(Z'\hat{\xi}(\theta))W(Z'\hat{\xi}(\theta))'$$

▶ When we have found $\hat{\theta}_{BLP}$ we can use this to update $W \to W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

BLP Alternatives

- BLP give us both a statistical estimator and an algorithm to obtain estimates.
- Plenty of other algorithms exist
    - We could solve for δ using the contraction mapping, using `fsolve` / Newton's Method / Guess and Check (not a good idea!).
    - We could try and consider a non-nested estimator for the BLP problem instead of solving for $\delta(\theta), \xi(\theta)$ we could let $\delta, \xi, \alpha, \beta$ be free parameters.
- We could think about different statistical estimators such as K-step GMM, Continuously Updating GMM, etc.

Dube Fox Su (2012)

$$\arg \min_{\theta_2} \quad \psi' \Omega^{-1} \psi \quad \text{s.t.}$$
$$\psi = \xi(\theta_2)' Z$$
$$\xi_{jt}(\theta) = \delta_j(\theta_2) - x_{jt}\beta - \alpha p_{jt} \qquad (1)$$
$$\log(S_{jt}) = \log(s_{jt}(\delta, \theta_2))$$

$$\arg \min_{\theta_2, \alpha, \beta, \xi, \psi} \quad \psi' \Omega^{-1} \psi \quad \text{s.t.}$$
$$\psi = \xi' Z \qquad (2)$$
$$\log(S_{jt}) = \log(s_{jt}(\xi, \theta_2, \alpha, \beta))$$

Comparing Approaches

- The original BLP paper and the DFS paper define different algorithms to produce the same statistical estimator.
  - The BLP algorithm is a nested fixed point (NFP) algorithm.
  - The DFS algorithm is a mathematical program with equilibrium constraints (MPEC).
  - The unknown parameters satisfy the same set of first-order conditions. (Not only asymptotically, but in finite sample).
  - $\hat{\theta}_{NFP} \approx \hat{\theta}_{MPEC}$ but for numerical differences in the optimization routine.
- Our choice of algorithm should mostly be about computational convenience.

BLP: NFP Advantages/Disadvantages

- ▶ Advantages
  - ▶ Concentrate out all of the linear in utility parameters (ξ, δ, β) so that we only search over Σ. When $\dim(\Sigma) = K$ is small (few dimensions of unobserved heterogeneity) this is a big advantage. For $K \leq 3$ this is my preferred approach.
  - ▶ When T (number of markets/periods) is large then you can exploit solving in parallel for δ market by market.
- ▶ Disadvantages
  - ▶ Small numerical errors in contraction can be amplified in the outer loop, $\rightarrow$ tolerance needs to be very tight.
  - ▶ Errors in numerical integration can also be amplified in the outer loop $\rightarrow$ must use a large number of draws/nodes.
  - ▶ Hardest part is working out the Jacobian via IFT.

$$D\delta_{\cdot t} = \begin{pmatrix} \dfrac{\partial \delta_{1t}}{\partial \theta_{21}} & \cdots & \dfrac{\partial \delta_{1t}}{\partial \theta_{2L}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \delta_{Jt}}{\partial \theta_{21}} & \cdots & \dfrac{\partial \delta_{Jt}}{\partial \theta_{2L}} \end{pmatrix} = - \begin{pmatrix} \dfrac{\partial s_{1t}}{\partial \delta_{1t}} & \cdots & \dfrac{\partial s_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial s_{Jt}}{\partial \delta_{1t}} & \cdots & \dfrac{\partial s_{Jt}}{\partial \delta_{Jt}} \end{pmatrix}^{-1} \begin{pmatrix} \dfrac{\partial s_{1t}}{\partial \theta_{21}} & \cdots & \dfrac{\partial s_{1t}}{\partial \theta_{2L}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial s_{Jt}}{\partial \theta_{21}} & \cdots & \dfrac{\partial s_{Jt}}{\partial \theta_{2L}} \end{pmatrix},$$

BLP: MPEC Advantages/Disadvantages

- Advantages
  - Problem scales better in $\dim(\Sigma)$.
  - Because all constraints hold at the optimum only: less impact of numerical error in tolerance or integration.
  - Derivatives are less complicated than $\frac{\partial \delta}{\partial \theta}$ (no IFT).
- Disadvantages
  - We are no longer concentrating out parameters, so there are a lot more of them! Storing the (Hessian) matrix of second derivatives can be difficult on memory.
  - We have to find the derivatives of the shares with respect to all of the parameters β, ξ, θ. (The other derivatives are pretty easy).
  - Parallelizing the derivatives is trickier than NFP case.

BLP Extensions: Demographics

▶ It is helpful to allow for interactions with consumer demographics (such as income).

$$\alpha_{it} = \overline{\alpha} + \sigma_p \nu_i + \pi_p y_{it}$$

▶ A few ways to do this:
  ▶ You could just use cross sectional variation in s_{jt} and $\overline{y}_t$ (mean or median income).
  ▶ Better: Draw y_{it} from a geographic specific income distribution. Draw ν_i from a general distribution of unobserved heterogeneity.

▶ Ex: Nevo (2000) Cereal demand sampled individual level D_i from geographic specific CPS data

▶ Joint distribution of income, income-squared, age, child at home.

$$\beta_i = \overline{\beta} + \Pi D_i + \sigma \nu_i$$

BLP Extensions: Panel Data

- with enough observations on the same product it is possible to include fixed effects

$$\delta_{jt}(\Sigma) = x_{jt}\beta - \alpha p_{jt} + \underbrace{\xi_{jt}}_{\xi_j + \xi_t + \Delta\xi_{jt}}$$

- What does ξ_j mean in this context?
- What would ξ_t mean in this context?
- $\Delta\xi_{jt}$ is now the structural error term, this changes our identification strategy a little.

Suppose we had additional data on behavior of individuals (in addition to aggregate market).

- ▶ Examples:
    - ▶ For some customers have answer to "Which car would you have purchased if the car you bought was not available?"
    - ▶ Demographic data on purchasers of a single brand.
    - ▶ Full individual demographic and choice data.

Extensions: Micro Data (Petrin 2002), (microBLP 2004)

▶ Previously we had moment conditions from orthogonality of structural error (ξ) and (X, Z) in order to form our GMM objective.

$$E[\xi_{jt}|x_{jt}, z_{jt}] = 0 \to E[\xi'[Z\,X]] = 0$$

▶ We can incorporate additional information using "micro-moments" or additional moment conditions to match the micro data.

  ▶ $Pr(\text{ i buys j } |y_i \in [0, \$20K]) = c_1$
  ▶ $Cov(d_i, s_{ijt}) = c_2$
  ▶ Construct an additional error term ζ_1, ζ_2 and interact that with instruments to form additional moment conditions.
  ▶ Econometrics get tricky when we have a different number of observations for $E[\zeta'[XZ]] = 0$ and $E[\xi'[XZ]] = 0$.

Alternative: Vertical Model (Bresnahan 1987)

- ▶ Imagine everyone agreed on the quality of the products offered for sale.
- ▶ The only thing people disagree on is willingness to pay for quality

$$U_{ij} = \overline{u} + \delta_j - \alpha_i p_j$$

- ▶ How do we estimate?
    - ▶ Sort goods from $p_1 < p_2 < p_3 \ldots < p_J$.
      It must be that $\delta_1 < \delta_2 < \ldots < \delta_J$. Why?
    - ▶ Normalize o.g. to 0 so that $0 > \delta_1 - \alpha_i p_1$ or $\alpha_i > \delta_1/p_1$.
    - ▶ $s_0 = F(\infty) - F(\frac{\delta_1}{p_1}) = 1 - F(\frac{\delta_1}{p_1})$ where $F(\cdot)$ is CDF of α_i.
    - ▶ In general choose j IFF:

$$\frac{\delta_{j+1} - \delta_j}{p_{j+1} - p_j} < \alpha_i < \frac{\delta_j - \delta_{j-1}}{p_j - p_{j-1}}$$

$$s_j = F\left(\frac{\delta_{j+1} - \delta_j}{p_{j+1} - p_j}\right) - F\left(\frac{\delta_j - \delta_{j-1}}{p_j - p_{j-1}}\right)$$

Alternative: Vertical Model (Bresnahan 1987)

Estimation

- ▶ Choose parameters θ of $F(\cdot)$ in order to best match s_j.
  - ▶ Can do MLE $\arg\max_\theta \sum_j \tilde{s}_j \log s_j(\theta)$.
  - ▶ Can do least squares $\sum_j (\tilde{s}_j - s_j(\theta))^2$.
  - ▶ Can do IV/GMM if I have an instrument for price.
    $\delta_j = x_j \beta + \xi_j$.
  - ▶ Extremely easy when $F \sim \exp(\lambda)$.
- ▶ What about elasticities?
  - ▶ When I change the price of j it can only affect (s_{j-1}, s_j, s_{j+1}).
  - ▶ We have set all of the other cross-price elasticities to be zero.
  - ▶ If a luxury car and a truck have similar prices, this can create strange substitution patterns.

Pure Characteristics Model: Berry Pakes (2001/2007)

$$u_{ij} = \delta_j + \sum_k \nu_{ik} x_{jk} + \xi_j + \underbrace{\sigma_i \epsilon_{ij}}_{\to 0}$$

- Can think of this like random coefficients model where we take the variance of ϵ to zero.
- Can think of this a vertical model, with vertical tastes over several characteristics.
    - PCs: everyone prefers more Mhz, more RAM, and more storage but differ in WTP.
    - Possible that there is no PC specific ϵ.
- Advantages
    - Logit error means there is always some substitution to all other goods.
    - Reality may be you only compete with a small number of competitors.
    - Allows for crowding in the product space.
- Disadvantage: no closed form for s_j, so estimation is extremely difficult.
- Minjae Song (Homotopy) and Che-Lin Su (MPCC) have made progress using two different approaches.

Adding Supply

Supply

- Economic theory gives us some additional powerful restrictions.
- We may want to impose $MR = MC$.
- Alternatively, we can ask – what is a good instrument for demand? something from another equation (ie: supply).

Bertrand Nash Pricing

Consider the problem of firm f which sets prices p_j for products in the set $\mathcal{J}_f$:

$$
\begin{aligned}
\arg \max_{p \in \mathcal{J}_f} \pi_f(\mathbf{p}) &= \sum_{j \in \mathcal{J}_f} (p_j - c_j) \cdot q_j(\mathbf{p}) \\
\to 0 &= q_j(\mathbf{p}) + \sum_{k \in \mathcal{J}_f} (p_k - c_k) \frac{\partial q_k}{\partial p_j}(\mathbf{p})
\end{aligned}
$$

We can re-write the FOC in matrix form:

$$
q(\mathbf{p}) = \Omega(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{mc})
$$

Bertrand Nash Pricing (cont)

It is helpful to define two $J_t \times J_t$ matrices. The first is an *ownership matrix*:

$$O_{(j,k)} = \left\{ \begin{array}{ll} 1 & \text{for } (j,k) \in \mathcal{J}_f \\ 0 & \text{for } (j,k) \notin \mathcal{J}_f \end{array} \right\}$$

And the second is the matrix of demand derivatives $\tilde{\Omega}(\mathbf{p})$ with entries $\tilde{\Omega}_{(j,k)}(\mathbf{p}) = \frac{\partial q_j}{\partial p_k}(\mathbf{p})$. We are mainly interested in the Hadamard (element-wise) product of the two matrices $\Omega = O \odot \tilde{\Omega}$.

$$\Omega_{(j,k)}(\mathbf{p}, \theta) = \left\{ \begin{array}{ll} -\frac{\partial q_j}{\partial p_k}(\mathbf{p}, \theta) & \text{for } (j,k) \in \mathcal{J}_f \\ 0 & \text{for } (j,k) \notin \mathcal{J}_f \end{array} \right\}$$

Recovering Marginal Costs

Recover implied markups/ marginal costs, and assume a functional form for $mc_{jt}(x_{jt}, w_{jt})$.

$$
\widehat{\mathbf{mc}}(\theta) = \mathbf{p} - \Omega(\mathbf{p}, \theta)^{-1} q(\mathbf{p}, \theta)
$$
$$
f(mc_{jt}) = x_{jt}\gamma_1 + w_{jt}\gamma_2 + \omega_{jt}
$$

Which we can solve for ω_{jt}:

$$
\omega_{jt} = f(\mathbf{p} - \Omega(\mathbf{p}, \theta)^{-1} q(\mathbf{p}, \theta)) - x_{jt}\gamma_1 - w_{jt}\gamma_2
$$

▶ $f(\cdot)$ is usually $\log(\cdot)$ or identity.
▶ I can use this to form additional moments: $E[\omega'_{jt} Z^s_{jt}] = 0$.
▶ I can just stack these up with the demand moments $E[\xi'_{jt} Z^d_{jt}] = 0$.
▶ Now I have $\dim(Z^d) + \dim(Z^s)$ moments altogether.
▶ This step is optional but can aid in identification (if you believe it).

Supply Side as Instruments

- ▶ Instruments for demand depend on exclusion restrictions
- ▶ Where do instruments come from? Something omitted that appears in another equation.

$$p_{jt} = c_{jt}(w_{jt}, x_{jt}) + \frac{s_{jt}(\mathbf{p_t})}{\left|\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}\right|} + \sum_{k \in \mathcal{J}_f} (p_k - c_k)\frac{\frac{\partial s_{kt}}{\partial p_{jt}}(\mathbf{p_t})}{\left|\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}\right|}$$

1. Exogenous regressors x_{jt}.
2. Cost shifters: w_{jt} (hard to find in practice), Hausman instruments.
3. Markup shifters: $\frac{s_{jt}(\mathbf{p_t})}{\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}}$

   (function of $(p_j, x_j, \xi_j, p_{-j}, x_{-j}, \xi_{-j})$).

Instruments and Identification

Instruments

- Recall the nested logit, where there are two separate endogeneity problems
  - Price: this is the familiar one!
  - Nonlinear characteristics σ this is the other one.
- We are doing nonlinear GMM: Start with $E[\xi_{jt}|x_{jt}, z_{jt}] = 0$ use $E[\xi'[ZX]] = 0$.
  - In practice this means that for valid instruments (x, z) any function $f(x, z)$ is also a valid instrument $E[\xi_{jt}f(x_{jt}, z_{jt})] = 0$.
  - We can use $x, x^2, x^3, \ldots$ or interactions $x \cdot z, x^2 \cdot z^2, \ldots$.
  - What is a reasonable choice of $f(\cdot)$?
  - Where does z come from?

Identifcation

- ▶ Once we have $\delta_{jt}(\theta)$ identification of linear parameters is pretty straightforward

$$\delta_{jt}(\theta) = x_{jt}\beta - \alpha p_{jt} + \xi_j + \xi_t + \Delta\xi_{jt}$$

- ▶ This is either basic linear IV or panel linear IV.
- ▶ How are σ taste parameters identified?
  - ▶ Consider increasing the price of j and measuring substitution to other products k, k' etc.
  - ▶ If sales of k increase with p_j and $(x_j^{(1)}, x_k^{(1)})$ are similar then we increase the σ that corresponds to $x^{(1)}$.
  - ▶ Price is the most obvious to vary, but sometimes this works for other characteristics (like distance).
  - ▶ Alternative: vary the set of products available to consumers by adding or removing an option.

▶ We can also impose the Bertrand FOC as a set of additional moments.

▶ First parametrize marginal cost

$$\ln mc_{jt} = \gamma_1 x_{jt} + \gamma_2 w_{jt} + \omega_{jt}$$

▶ helpful to constrain MC to be positive always.

▶ Note that for any vector of prices p and demand parameters θ we can recover a unique vector of marginal costs (by solving the system of linear equations).

▶ Imposing the supply side only helps if we have information about the marginal costs / production function that we would like to impose

▶ Imposing these restrictions is helpful in constraining markups (so that implied MC are always positive, etc.).

▶ Misspecified functional forms for costs can cause problems!

Instruments

- Common choices are average characteristics of other products in the same market $h(x_{-j,t})$. BLP instruments
  - Same firm $z_{1jt} = \overline{x}_{-j_f,t} = \frac{1}{|F_j|} \sum_{k \in \mathcal{F}_j} x_{kt} - \frac{1}{|F_j|} x_{jt}$.
  - Other firms $z_{2jt} = \overline{x}_{\cdot t} - \overline{x}_{-j_f,t} - \frac{1}{J} x_{jt}$.
  - Plus regressors $(1, x_{jt})$.
  - Plus higher order interactions
- Technically linearly independent for large (finite) J, but becoming highly correlated.
  - Can still exploit variation in number of products per market or number of products per firm.
- Correlated moments $\rightarrow$ "many instruments".
  - May be inclined to "fix" correlation in instrument matrix directly.

Armstrong (2016): Weak Instruments?

Consider the limit as $J \to \infty$

$$\frac{s_{jt}(\mathbf{p_t})}{\left|\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}\right|} = \frac{1}{\alpha}\frac{1}{1 - s_{jt}} \to \frac{1}{\alpha}$$

▶ Hard to use markup shifting instruments to instrument for a constant.

▶ How close to the constant do we get in practice?

▶ Average of x_{-j} seems like an especially poor choice. Why?

▶ Shows there may still be some power in: products per market, products per firm.

▶ Convergence to constant extends to mixed logits (see Gabaix and Laibson 2004).

▶ Evidence that you really need cost shifters.

Optimal Instruments

How to construct optimal instruments in form of Chamberlain (1987)

$$E\left[\frac{\partial \xi_{jt}}{\partial \theta}|X_t, w_{jt}\right] = \left[\beta, E\left[\frac{\partial \xi_{jt}}{\partial \alpha}|X_t, w_{jt}\right], E[\frac{\partial \xi_{jt}}{\partial \sigma}|X_t, w_{jt}]\right]$$

Some challenges:

1. p_{jt} depends on X_t, w_t, ξ_t in a highly nonlinear way (no explicit solution!).

2. $E[\frac{\partial \xi_{jt}}{\partial \sigma}|X_t, w_t] = E[[\frac{\partial \mathbf{s_t}}{\partial \delta_t}]^{-1}[\frac{\partial \mathbf{s_t}}{\partial \sigma}]|X_t, w_t]$ (not conditioned on endogenous p!)

"Feasible" Recipe:

1. Fix $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ and draw ξ_t from empirical density

2. Solve fixed point equation for $\hat{p_{jt}}$

3. Compute necessary Jacobian

4. Average over all values of ξ_t. (Lazy approach: use only $\xi = 0$).

Optimal Instruments

- Since any $f(x, z)$ satisfies our orthogonality condition, we can try to choose $f(x, z)$ as a basis to approximate optimal instruments.
- This is challenging in practice – and in fact suffers from a curse of dimensionality.
- This is frequently given as a rationale behind higher order x's.
- When the dimension of x is low – this may still be feasible. $(K \leq 3)$.

- Optimal instruments are easier to work out if $p = mc$.

$$c = p + \underbrace{\Delta^{-1}}_{\to 0} s = X\gamma_1 + W\gamma_2 + \omega$$

- Linear cost function means linear reduced-form price function.

$$
\begin{aligned}
E\left[\frac{\partial \xi_{jt}}{\partial \alpha} | z_t\right] &= E[p_{jt}|z_t] = x_{jt}\gamma_1 + w_{jt}\gamma_2 \\
E\left[\frac{\partial \omega_{jt}}{\partial \alpha} | z_t\right] &= 0, \quad E\left[\frac{\partial \omega_{jt}}{\partial \sigma} | z_t\right] = 0 \\
E\left[\frac{\partial \xi_{jt}}{\partial \sigma} | z_t\right] &= E\left[\frac{\partial \delta_{jt}}{\partial \sigma} | z_t\right]
\end{aligned}
$$

- If we are worried about endogenous oligopoly markups is this a reasonable idea?

Optimal Instruments: Reynaert Verboven (2014)

Table 2: Bias and Efficiency with Imperfect Competition

		Single Equation GMM								
		g_{jt}^1			g_{jt}^2			g_{jt}^3		
	True	Bias	St Err	RMSE	Bias	St Err	RMSE	Bias	St Err	RMSE
β^0	2	-0.127	0.899	0.907	-0.155	0.799	0.814	-0.070	0.514	0.519
β^1	2	-0.068	0.899	0.901	0.089	0.766	0.770	-0.001	0.398	0.398
α	-2	0.006	0.052	0.052	0.010	0.049	0.050	0.010	0.043	0.044
σ^1	1	-0.162	0.634	0.654	-0.147	0.537	0.556	-0.016	0.229	0.229
		Joint Equation GMM								
		g_{jt}^1			g_{jt}^2			g_{jt}^3		
	True	Bias	St Err	RMSE	Bias	St Err	RMSE	Bias	St Err	RMSE
β^0	2	-0.095	0.714	0.720	-0.103	0.677	0.685	0.005	0.459	0.459
β^1	2	0.089	0.669	0.675	0.098	0.621	0.628	-0.009	0.312	0.312
α	-2	0.001	0.047	0.047	0.002	0.046	0.046	-0.001	0.043	0.043
σ^1	1	-0.116	0.462	0.476	-0.110	0.418	0.432	0.003	0.133	0.133

Bias, standard errors (St Err) and root mean squared errors (RMSE) are computed from 1000 Monte Carlo replications. Estimates are based on the MPEC algorithm and Sparse Grid integration. The instruments g_{jt}^1, g_{jt}^2 and g_{jt}^3 are defined in section 2.4 and 2.5.

Differentiation Instruments: Gandhi Houde (2016)

▶ Also need instruments for the Σ or σ random coefficient parameters.

▶ Instead of average of other characteristics $h(x) = \frac{1}{J-1} \sum_{k \neq j} x_k$, can transform as distance to x_j.

$$d_{jt}^k = x_k - x_j$$

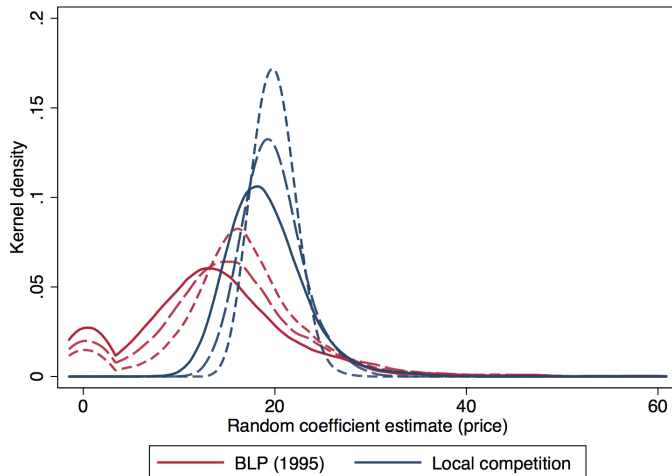▶ And use this transformed to construct two kinds of IV (Squared distance, and count of local competitors)

$$DIV_1 = \qquad \sum_{j \in F} d_{jt}^2, \qquad \sum_{j \notin F} d_{jt}^2$$

$$DIV_2 = \sum_{j \in F} I[d_{jt} < c] \qquad \sum_{j \notin F} I[d_{jt} < c]$$

▶ They choose c to correspond to one standard deviation of x across markets.

Differentiation Instruments: Gandhi Houde (2016)

Figure 4: Distribution of parameter estimates in small and large samples

Sample size: Solid = 500, Long dash = 1,000, Dash = 2,500.

Differentiation Instruments: Gandhi Houde (2016)

Table 5: Monte-Carlo simulations with endogenous prices.

		IV: Sum of charact.		IV: Local competitors	
		w/o cost	w/ cost	w/o cost	w/ cost
β_p					
	Average	0.46	1.08	1.00	1.02
	RMSE	2.19	1.32	0.22	0.18
σ_p					
	Average	13.24	17.47	19.10	19.68
	RMSE	10.84	7.95	3.93	1.51

(a) Market versus Differentiation IVs

		IV: Sum of charact.		IV: Local competitors	
		Market IV	Opt. IV	Diff. IV	Opt. IV
β_p					
	Average	0.46	1.29	1.00	1.16
	RMSE	2.19	0.93	0.22	0.45
σ_p					
	Average	13.24	16.61	19.10	17.28
	RMSE	10.84	28.23	3.93	19.07

(b) Optimal IV approximation without cost shifter

Sample size: $\bar{n}_t \times T = 500$. Number of Monte-Carlo simulations = 1,000. True parameter value: $\sigma_p = 20$, $\beta_p = 1$.