Best Practices for Demand Estimation with `pyblp`

Chris Conlon and Jeff Gormaker

NYU Stern and NY Fed

BAM!- December 2018

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.

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. $E[\frac{\partial \xi_{jt}}{\partial \sigma}|X_t, w_t] = E[[\frac{\partial \mathbf{s_t}}{\partial \delta_\mathbf{t}}]^{-1}[\frac{\partial \mathbf{s_t}}{\partial \sigma}]|X_t, w_t]$ (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: Reynaert Verboven (2014)

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

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

- 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?

Table 2: Bias and Efficiency with Imperfect Competition

		Single Equation GMM								
		g^1_{jt}			g^2_{jt}			g^3_{jt}		
	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^1_{jt}			g^2_{jt}			g^3_{jt}		
	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^1_{jt}, g^2_{jt}.and g^3_{jt} 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.

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

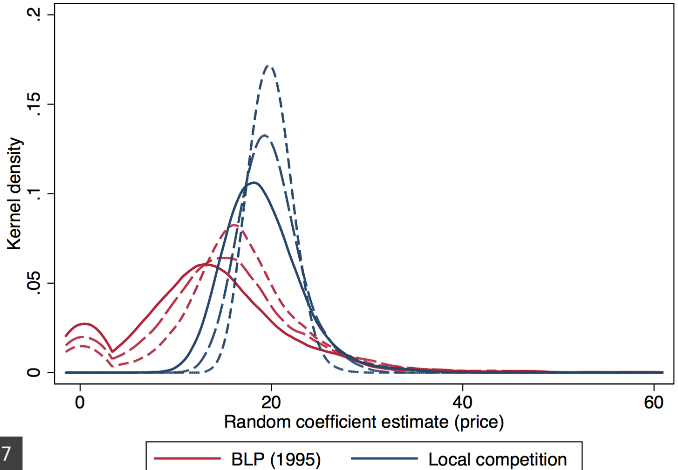
- And use this transformed to construct two kinds of IV (Squared distance, and count of local competitors)

$$\begin{aligned} DIV_1 = \quad & \textstyle\sum_{j\in F} d_{jt}^2, && \sum_{j\notin F} d_{jt}^2 \\ DIV_2 = \quad & \textstyle\sum_{j\in F} I[d_{jt} < c] && \sum_{j\notin F} I[d_{jt} < c] \end{aligned}$$

- 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



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

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

An Ongoing Project...

What do we have so far?

- Available on `PyPI`

  `pip install pyblp`
- Extensive documentation: `https://pyblp.readthedocs.io/en/stable/`
- Long list of features
- 6k downloads: who are these people?

Some setup

We can break up the parameter space into three parts:

- θ_1: linear exogenous demand parameters,
- θ_2: nonlinear endogenous parameters including price and random coefficients
- θ_3: linear exogenous supply parameters.

The basic setup

(a) For each market t: solve $\mathcal{S}_{jt} = s_{jt}(\delta_{\cdot t}, \theta_2)$ for $\widehat{\delta}_{\cdot t}(\theta_2)$.

(b) For each market t: use $\widehat{\delta}_{\cdot t}(\theta_2)$ to construct $\eta_{\cdot t}(\mathbf{q_t}, \mathbf{p_t}, \widehat{\delta}_{\cdot t}(\theta_2), \theta_2)$

(c) For each market t: Recover $\widehat{mc}_{jt}(\widehat{\delta}_{\cdot t}(\theta_2), \theta_2) = p_{jt} - \eta_{jt}(\widehat{\delta}_{\cdot t}(\theta_2), \theta_2)$

(d) Stack up $\widehat{\delta}_{\cdot t}(\theta_2)$ and $\widehat{mc}_{jt}(\widehat{\delta}_{\cdot t}(\theta_2), \theta_2)$ and use linear IV-GMM to recover $[\widehat{\theta}_1(\theta_2), \widehat{\theta}_3(\theta_2)]$ following the recipe in Appendix

(e) Construct the residuals:

$$\widehat{\xi}_{jt}(\theta_2) = \widehat{\delta}_{jt}(\theta_2) - x_{jt}\widehat{\beta}(\theta_2) + \alpha p_{jt}$$
$$\widehat{\omega}_{jt}(\theta_2) = \widehat{mc}_{jt}(\theta_2) - [x_{jt}\, w_{jt}]\,\widehat{\gamma}(\theta_2)$$

(f) Construct sample moments

$$g_n^D(\theta_2) = \frac{1}{N}\sum_{jt} Z_{jt}^{D\prime}\widehat{\xi}_{jt}(\theta_2)$$
$$g_n^S(\theta_2) = \frac{1}{N}\sum_{jt} Z_{jt}^{S\prime}\widehat{\omega}_{jt}(\theta_2)$$

(g) Construct GMM objective $Q_n(\theta_2) = \begin{bmatrix} g_n^d(\theta_2) \\ g_n^s(\theta_2) \end{bmatrix}' W \begin{bmatrix} g_n^d(\theta_2) \\ g_n^s(\theta_2) \end{bmatrix}$

Additional Details

Some different definitions:

$$\begin{aligned} y_{jt}^{D} &:= \widehat{\delta}_{jt}(\theta_2) + \alpha p_{jt} && = x_{jt}'\beta + \xi_t =: x_{jt}^{D\prime}\beta + \xi_{jt} \\ y_{jt}^{S} &:= \widehat{mc}_{jt}(\theta_2) && = (x_{jt}\ w_{jt})'\gamma + \omega_t =: x_{jt}^{S\prime}\gamma + \omega_{jt} \end{aligned} \tag{1}$$

Stacking the system across observations yields:[1]

$$\underbrace{\begin{bmatrix} y_D \\ y_S \end{bmatrix}}_{2N\times 1} = \underbrace{\begin{bmatrix} X_D & 0 \\ 0 & X_S \end{bmatrix}}_{2N\times(K_1+K_3)} \underbrace{\begin{bmatrix} \beta \\ \gamma \end{bmatrix}}_{(K_1+K_3)\times 1} + \underbrace{\begin{bmatrix} \xi \\ \omega \end{bmatrix}}_{2N\times 1} \tag{2}$$

[1] Note: we cannot perform independent regressions unless we are willing to assume that $Cov(\xi_{jt}, \omega_{jt}) = 0$.

#1: Solving the Contraction

BLP also propose a fixed-point approach to solve the $J_t \times J_t$ system of equations for shares. They show that the following is a contraction mapping $f(\delta) = \delta$:

$$f : \delta_{\cdot t}^{h+1} \leftarrow \delta_{\cdot t}^{h} + \ln \mathcal{S}_{\cdot t} - \ln \mathbf{s_{\cdot t}}(\delta_{\cdot t}^{h}, \theta_2) \tag{3}$$

- This kind of contraction mapping is linearly convergent where the rate of convergence is proportional to $\frac{L(\theta_2)}{1-L(\theta_2)}$ where $L(\theta_2)$ is the Lipschitz constant.
- Because (3) is a contraction, we know that $L(\theta_2) < 1$.
- DFS2012 show that for the BLP contraction the Lipschitz constant is defined as $L(\theta_2) = \max_{\delta \in \Delta} \left\| \mathbf{I}_{J_t} - \frac{\partial \log \mathbf{s_{\cdot t}}}{\partial \delta_{\cdot t}}(\delta_{\cdot t}, \theta_2) \right\|_{\infty}$.

#1: Accelerating the Contraction (Newton's Method)

$$\delta_{\cdot t}^{h+1} \leftarrow \delta_{\cdot t}^{h} - \lambda J_{\mathbf{s}}^{-1}(\delta_{\cdot t}^{h}, \theta_2) \cdot \mathbf{s_t}(\delta_{\cdot t}^{h}, \theta_2)$$

- Each Newton-Raphson iteration would require computation of:
  - J_t vector of marketshares $\mathbf{s_t}(\delta_{\cdot t}^{h}, \theta_2)$, the $J_t \times J_t$
  - Jacobian matrix $J_{\mathbf{s}}(\delta_{\cdot t}^{h}, \theta_2) = \frac{\partial \mathbf{s_{\cdot t}}}{\partial \delta_{\cdot t}}(\delta_{\cdot t}^{h}, \theta_2)$
  - as well as its inverse $J_{\mathbf{s}}^{-1}(\delta_{\cdot t}^{h})$.
- Inverse Jacobian can be costly when J_t is large (and requires integration).
- Can speed up using *Anderson* or *Aitken Acceleration*.

Most of these methods use information from mutliple iterations $(\delta^h, \delta^{h+1}, \delta^{h+2}, f(\delta^h), f(f(\delta^h)))$ to approximate $J_{\mathbf{s}}$ or $J_{\mathbf{s}}^{-1}$:

$$\begin{aligned} \delta_{\cdot t}^{h+1} &= \delta_{\cdot t}^{h} - 2\alpha^h r^h + (\alpha^h)^2 v^h, \quad \alpha^h = \frac{(v^h)' r^h}{(v^h)' v^h} \\ r^h &= f(\delta_{\cdot t}^h) - \delta_{\cdot t}^h, \quad v^h = f(f(\delta_{\cdot t}^h)) - 2f(\delta_{\cdot t}^h) + \delta_{\cdot t}^h \end{aligned} \tag{4}$$

- This particular algorithm is known as SQUAREM used in biostats for EM algorithms.
- Applied to BLP by Reynaerts, Varadhan, Nash (2012).
- Iterations are more costly but much more accurate (almost a Newton step).
- Speedup is 2-12x.

#2: High Dimensional Fixed Effects

- Suppose I want to incorporate store-upc and store-week FE using Nielsen Data.
  - Around 500 weeks since 2006.
  - Around 3000+ UPCs in a category like distilled spirits or breakfast cereal.
  - Can easily find ourselves estimating 50,000+ fixed effects in a single dimension and several thousand in the other.

#2: High Dimensional Fixed Effects

There are several differencing algorithms for removing the fixed effects. For simplicity let's assume there are two dimensions of fixed effects N and T where $N >> T$:

$$\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t}$$
$$\tilde{x}_{it} = x_{it} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t}$$

- Could do *iterative demeaning*: easy if $Cov(\bar{x}_{\cdot t}, \bar{x}_{i\cdot}) = 0$. Otherwise hard.
- LSDV requires inverting the $(N+T) \times (N+T)$ matrix which can be difficult to impossible.
- `reghdfe` like Correia (2016) does iterative projection of x on y and never inverts the matrix but FE always depend on both (y, x).

#2: High Dimensional Fixed Effects: Somaini Wolak (2016)

Adapt Method of Somaini Wolak (2016):

- Use FLW theorem and only invert the $(T \times T)$ matrix.
- The key is that we can demean x once and never again.
- Speedup is big. Memory usage is limited.
- Can get into trouble when both dimensions are very large.

Recall the multi-product Bertrand FOCs:

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

It is helpful to define the matrix Ω with entries:

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

We can re-write the FOC in matrix form:

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

- Can we iterate on the price relation until we converge to a new equilibrium?

$$\mathbf{p} \leftarrow \widehat{\mathbf{mc}} - \Omega(\mathbf{p})^{-1} q(\mathbf{p})$$

- While tempting, this doesn't work. (It is not a contraction).
- There is a modification that is a contraction for logit type models.
- You can always get lucky(!)

- For the logit (and variants) we can factor $\frac{\partial q_j}{\partial p_k}$ into two parts.

$$\Omega_{jk}(\mathbf{p}) = \underbrace{\alpha \cdot I[j = k] \cdot s_j(\mathbf{p})}_{\Lambda(\mathbf{p})} - \underbrace{\alpha \cdot s_j(\mathbf{p}) s_k(\mathbf{p})}_{\Gamma(\mathbf{p})}$$

- $\Gamma(\mathbf{p})$ and $\Lambda(\mathbf{p})$ are $J \times J$ matrices and $\Lambda(\mathbf{p})$ is diagonal and (j, k) is nonzero in $\Gamma(\mathbf{p})$ only if (j, k) share an owner.
- After factoring we can rescale by $\Lambda^{-1}(\mathbf{p})$

$$(\mathbf{p} - \mathbf{mc}) \leftarrow \Lambda^{-1}(\mathbf{p}) \cdot \Gamma(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{mc}) - \Lambda^{-1}(\mathbf{p}) \cdot s(\mathbf{p})$$

- This alternative fixed point is in fact a contraction.
- Moreover the rate of convergence is generally fast and stable (much more than Gauss-Seidel or Gauss-Jacobi).

Chamberlain (1987) tells us the optimal instruments for this supply-demand system of $G\Omega^{-1}$ where for a given observation n,

$$G_n := \underbrace{\begin{bmatrix} \frac{\partial\xi}{\partial\beta} & \frac{\partial\omega}{\partial\beta} \\ \frac{\partial\xi}{\partial\alpha} & \frac{\partial\omega}{\partial\alpha} \\ \frac{\partial\xi}{\partial\sigma} & \frac{\partial\omega}{\partial\sigma} \\ \frac{\partial\xi}{\partial\gamma} & \frac{\partial\omega}{\partial\gamma} \end{bmatrix}_n}_{(K_1+K_2+K_3)\times 2} = \begin{bmatrix} -X & 0 \\ \xi_\alpha & \omega_\alpha \\ \xi_\sigma & \omega_\sigma \\ 0 & -X \\ 0 & -W \end{bmatrix}_n \qquad \Omega := \underbrace{\begin{bmatrix} V_\xi^2 & V_{\xi\omega} \\ V_{\xi\omega} & V_\omega^2 \end{bmatrix}}_{2\times 2}$$

#4: Optimal Instruments

$$G_n\Omega^{-1} = \frac{1}{V_\xi^2 V_\omega^2 - (V_{\xi\omega})^2} \times \begin{bmatrix} -V_\omega^2 X & V_{\xi\omega} X \\ V_\omega^2 \xi_\alpha - V_{\xi\omega}\omega_\alpha & V_\xi^2 \omega_\alpha - V_{\xi\omega}\xi_\alpha \\ V_\omega^2 \xi_\sigma - V_{\xi\omega}\omega_\sigma & V_\xi^2 \omega_\sigma - V_{\xi\omega}\xi_\sigma \\ V_{\xi\omega} X & -V_\xi^2 X \\ V_{\xi\omega} W & -V_\xi^2 W \end{bmatrix}_n$$

Clearly rows 1 and 4 are co-linear.

$$(G_n\Omega^{-1}) \circ \Theta = \frac{1}{v_\xi^2 v_\omega^2 - (v_{\xi\omega})^2} \times \begin{bmatrix} -v_\omega^2 x & 0 \\ v_\omega^2 \xi_\alpha - v_{\xi\omega}\omega_\alpha & v_\xi^2 \omega_\alpha - v_{\xi\omega}\xi_\alpha \\ v_\omega^2 \xi_\sigma - v_{\xi\omega}\omega_\sigma & v_\xi^2 \omega_\sigma - v_{\xi\omega}\xi_\sigma \\ 0 & -v_\xi^2 x \\ v_{\xi\omega} w & -v_\xi^2 w \end{bmatrix}_n$$

Now we can partition our instrument set by column into "demand" and "supply" instruments as

$$z_{nD} := (G_n\Omega^{-1} \circ \Theta)_{\cdot 1}$$
$$z_{nS} := (G_n\Omega^{-1} \circ \Theta)_{\cdot 2}$$

Aside: What does Supply tell us about Demand?

$$\begin{aligned} \partial\alpha : V_{\omega}^{2}\xi_{\alpha} - V_{\xi\omega}\omega_{\alpha} \quad & V_{\xi}^{2}\omega_{\alpha} - V_{\xi\omega}\xi_{\alpha} \\ \partial\sigma : V_{\omega}^{2}\xi_{\sigma} - V_{\xi\omega}\omega_{\sigma} \quad & V_{\xi}^{2}\omega_{\sigma} - V_{\xi\omega}\xi_{\sigma} \end{aligned}$$

- Under optimal IV these are overidentifying restrictions
- Maybe cases where one part of these instruments is trivial.

Demo

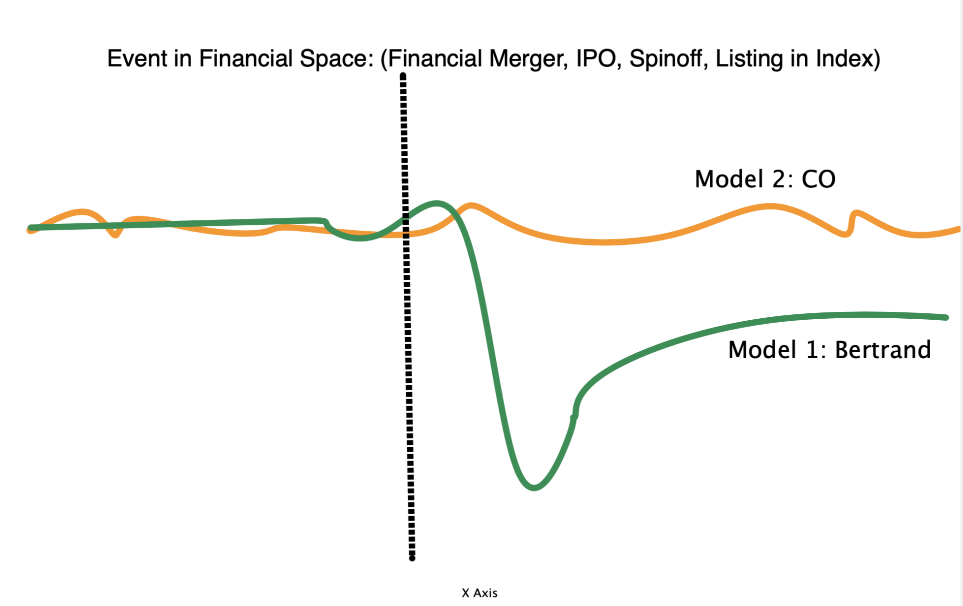
Testing for Conduct

Testing for Conduct

- General challenge is that if you give a me a vector of prices, I can always deliver a vector of marginal costs which rationalizes
- Estimation is either simultaneous supply and demand or sequential estimating demand first and then estimating supply separately. We can label the two sets of moment restrictions:

$$g^d(\theta_1, \theta_2) : E[\xi_{jt} \times [x_{jt}, z_{jt}, \eta^d_{jt}]] = 0$$
$$g^s(\theta_3, \kappa) : E[\omega_{jt} \times [x_{jt}, z_{jt}, \eta^s_{jt}]] = 0$$

- We could estimate simultaneously under two different conduct assumptions and compare GMM objectives.
  - But the weighting matrix changes, and the scale of y^s may not be the same.
  - Usually we have to test pairwise A rejects B and B rejects A.


Event in Financial Space: (Financial Merger, IPO, Spinoff, Listing in Index)
Model 2: CO
Model 1: Bertrand
X Axis

Another way to estimate using moment restrictions

- Re-weight my data so that my moments hold exactly

$$\sum_i \pi_i g(x_i, \theta) = 0 \quad \sum_i \pi_i = 1 \quad \pi_i \geq 0$$

- Choose the set of weights as close as possible to $\frac{1}{N}$ empirical weights.

$$l_{EL}(\pi, \theta) = \sum_i \log \pi_i \quad l_{CUE}(\pi, \theta) = -\sum_i \pi_i^2$$

Empirical Likelihood

Another way to estimate using moment restrictions

- Re-weight my data so that my moments hold exactly

$$\sum_i \pi_i g(x_i, \theta) = 0 \quad \sum_i \pi_i = 1 \quad \pi_i \geq 0$$

- Choose the set of weights as close as possible to $\frac{1}{N}$ empirical weights.

$$l_{EL}(\pi, \theta) = \sum_i \log \pi_i \quad l_{CUE}(\pi, \theta) = -\sum_i \pi_i^2$$

Full Problem

$$EL(\theta,\kappa) = \max_{\theta_1,\theta_2,\theta_3,\pi} \sum_{jt}^{N} \log \pi_{jt} \quad \text{s.t.} \quad \begin{aligned} \sum_{jt}^{N} \pi_{jt}\omega_{jt}(\kappa) \times [x_{jt}, z_{jt}, \eta^s_{jt}]] &= 0 \\ \sum_{jt}^{N} \pi_{jt}\xi_{jt} \times [x_{jt}, z_{jt}, \eta^d_{jt}]] &= 0 \\ s_{jt}(\theta) &= S_{jt} \end{aligned}$$

Restricted Problem

$$EL(\theta_3,\kappa) = \max_{\theta_3,\pi} \sum_{jt}^{N} \log \pi_{jt} \quad \text{s.t.} \quad \begin{aligned} \sum_{jt}^{N} \pi_{jt}\omega_{jt}(\kappa) \times [x_{jt}, z_{jt}, \eta^s_{jt}]] &= 0 \\ \sum_{jt}^{N} \pi_{jt}\widehat{\xi}_{jt} \times [x_{jt}, z_{jt}, \eta^d_{jt}]] &= 0 \end{aligned}$$

We cheat a little bit and fix the demand parameters $\widehat{\theta}_1, \widehat{\theta}_2$ but still utilize the demand moments. This lets us estimate sequentially but still utilize the restrictions from the

Likelihood Ratio Test

It is clear that we can just compare using a likelihood ratio test because we have the same number of parameters and restrictions in each model:

$$2 * \left[EL(\widehat{\theta}_3, \kappa) - EL(\tilde{\theta}_3, \kappa = 0)\right] \sim \chi^2_0$$

Overidentifying Restrictions

Add another instrument so that $\eta \rightarrow \widetilde{\eta}$.

$$2 * \left[EL(\widehat{\theta}_3, \kappa = 0, \eta) - EL(\tilde{\theta}_3, \kappa = 0, \widetilde{\eta})\right] \sim \chi^2_1$$

How to apply?

1. Cost shifters for other products: z_{kt} for p_{jt}. (Cost of Rice for Corn Flakes, Cost of Corn for Rice Krispies).
2. Strongest instrument should be the markup shifter $\kappa_{fg} \sum_{k \in \mathcal{J}_g} (p_k - c_k) \cdot D_{jk}$ or $\sum_{k \in \mathcal{J}_g} (p_k - c_k) \cdot D_{jk}$. Is this valid?
3. What belongs in $x_{jt}\beta$: FE or not? We can also consider higher order polynomials for ω moments.
4. The validity for many of these tests will often depend on how well we explain marginal costs (lower variance of ω) rather than specific exclusion restrictions. I think this makes me want to test the overidentifying restriction directly rather than goodness of fit type GMM or LR objective tests.
5. Test Bertrand against an arbitrary alternative by testing exclusion of diversion weighted markups to g's products. Under the null of Bertrand $\lambda = 0$. We can include this measure for each competitor's product:

$$\lambda_{jg} \sum_{k \in \mathcal{J}_g} (p_k - c_k) \cdot D_{jk}$$