Instruments and Identification

Chris Conlon

September 22, 2020

Grad IO

Parametric Identifcation

- Once we have $D_{jt}^{-1}(\mathcal{S}_t, \widetilde{\theta}_2) \equiv \delta_{jt}(\theta_2)$ identification of linear parameters is pretty straightforward

$$D_{jt}^{-1}(\mathcal{S}_t, \widetilde{\theta}_2) = x_{jt}\beta - \alpha \cdot p_{jt} + \xi_j + \xi_t + \Delta\xi_{jt}$$

- This is either basic linear IV or panel linear IV and we need instruments for p_{jt}
- The $\widetilde{\theta}_2$ parameters governing the change of variables require nonlinear IV
- Remember $\theta_2 = [\widetilde{\theta}_2, \alpha]$.

Exclusion Restrictions

$$\begin{aligned} \delta_{jt}(\mathcal{S}_t, \widetilde{\theta}_2) &= h_d\left(x_{jt}, v_{jt}; \theta_1\right) - \alpha p_{jt} + \xi_{jt} \\ f(p_{jt} - \eta_{jt}(\theta_2, \mathbf{p}, \mathbf{s})) &= h_d\left(x_{jt}, w_{jt}; \theta_3\right) + \omega_{jt} \end{aligned}$$

The first place to look for exclusion restrictions/instruments:

- Something in another equation!
- v_j shifts demand but not supply
- w_j shifts supply but not demand
- If it doesn't shift either is it really relevant?

Markup Shifters

The equilibrium markup is a function of everything! $\eta_{jt}(\mathbf{p}, \mathbf{s}, \xi_t, \omega_t, x_t, w_t, v_t, \theta_2)$:

- It is literally endogenous (depends on error terms)!
- But lots of potential instruments beyond excluded v_t or w_t.
- Also v_{-j} and w_{-j} and x_{-j} (these don't depend on ξ_{jt}, ω_{jt})
- Not p_{-j} or ξ_{-j}, (these depend on ξ_{jt}, ω_{jt}).
- The idea is that these instruments shift or rotate the marginal revenue curve.
- What is a good choice of $f(x_{-j})$? etc.

What are we instrumenting for?

- Recall the nested logit, where there are two separate endogeneity problems
  - Endogenous markups η_{jt} (link S+D)
  - Nonlinear characteristics σ on $\ln s_{j|gt}$ this is the other one.
- Nonlinear parameters $\widetilde{\theta}_2$.
  - 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. In which dimension are close substitutes "more similar".

Instruments

- We are doing nonlinear GMM: Start with $E[\xi_{jt}|x_{jt}, w_{jt}] = 0$ use $E[\xi'_{jt} Z^D_{jt}] = 0$ with $Z^D_{jt} = [x_{jt},\ z_{jt}]$.
  - In practice this means that for valid instruments (x, w) any function $f(x, w)$ is also a valid instrument $E[\xi_{jt} f(x_{jt}, w_{jt})] = 0$.
  - We can use $x, x^2, x^3, \ldots$ or interactions $x \cdot w, x^2 \cdot w^2, \ldots$.
  - What is a reasonable choice of $f(\cdot)$?
  - Where does w come from?

BLP Instruments

- Common choices are average characteristics of other products in the same market $f(x_{-j,t})$. 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).
- Suggests that you really need cost shifters.

Differentiation Instruments: Gandhi Houde (2019)

- Also need instruments for the random coefficient parameters $\widetilde{\theta}_2$.
- Instead of average of other characteristics $f(x) = \frac{1}{J-1} \sum_{k \neq j} x_k$, can transform as distance to x_j.

$$d_{jkt} = x_{kt} - x_{jt}$$

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

$$\begin{aligned} z_{jt}^{\textsf{quad}} &= \quad \textstyle\sum_{k \in F} d_{jkt}^2, \qquad \sum_{k \notin F} d_{jkt}^2 \\ z_{jt}^{\textsf{local}} &= \quad \textstyle\sum_{k \in F} I[d_{jkt} < c] \qquad \sum_{k \notin F} I[d_{jkt} < c] \end{aligned}$$

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

Optimal Instruments (Chamberlain 1987)

Chamberlain (1987) asks how can we choose $f(z_i)$ to obtain the semi-parametric efficiency bound with conditional moment restrictions:

$$E[g(z_i,\theta)|z_i] = 0 \Rightarrow E[g(z_i,\theta)\cdot f(z_i)] = 0$$

Recall that the asymptotic GMM variance depends on $(D'\,\Omega^{-1}D\,)$

The answer is to choose instruments related to the (expected) Jacobian of moment conditions w.r.t θ. The true Jacobian at θ_0 is infeasible:

$$D = E\left[\frac{\partial g(z_i,\theta)}{\partial\theta}|z_i\right]$$

Optimal Instruments (Chamberlain 1987)

Consider the simplest IV problem:

$$
\begin{aligned}
y_i &= \beta x_i + \gamma v_i + u_i \quad \text{with} \quad E[u_i | v_i, z_i] = 0 \\
u_i &= (y_i - \beta x_i - \gamma v_i) \\
g(x_i, v_i, z_i) &= (y_i - \beta x_i - \gamma v_i) \cdot [v_i,\ z_i]
\end{aligned}
$$

Which gives:

$$
\begin{aligned}
E\left[\frac{\partial g(x_i, v_i, z_i, \theta)}{\partial \gamma} \mid v_i, z_i\right] &= v_i \\
E\left[\frac{\partial g(x_i, v_i, z_i, \theta)}{\partial \beta} \mid v_i, z_i\right] &= E\left[x_i \mid v_i, z_i\right]
\end{aligned}
$$

We can't just use x_i (bc endogenous!), but you can also see where 2SLS comes from...

Optimal IV: BLP

Recall the GMM moment conditions are given by $E[\xi_{jt}|Z_{jt}^D] = 0$ and $E[\omega_{jt}|Z_{jt}^S] = 0$ and the asymptotic GMM variance depends on $(D' \, \Omega^{-1} D \,)$ where the expressions are given below:

$$D = E\left[\left(\frac{\partial \xi_{jt}}{\partial \theta}, \frac{\partial \omega_{jt}}{\partial \theta}\right) |\mathbf{Z_t}\right], \quad \Omega = E\left[\begin{pmatrix} \xi_{jt} \\ \omega_{jt} \end{pmatrix} \begin{pmatrix} \xi_{jt} \; \omega_{jt} \end{pmatrix} |\mathbf{Z_t}\right].$$

Chamberlain (1987) showed that the approximation to the optimal instruments are given by the expected Jacobian contribution for each observation (j,t): $E[D_{jt}(\mathbf{Z_t})\Omega_{jt}^{-1}|\mathbf{Z_t}]$.

Optimal Instruments (Newey 1990)

From previous slide, nothing says that $E\left[x_i \mid v_i, z_i\right]$ needs to be linear!

- Since any $f(x,z)$ satisfies our orthogonality condition, we can try to choose $f(x,z)$ as a basis to approximate optimal instruments.
- Why? Well affine tranformations of instruments are still valid, and we span the same vector space!
- We are essentially relying on a non-parametric regression that we never run (but could!)
  - 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. $(K \leq 3)$.
  - But recent improvements in sieves, LASSO, non-parametric regression are encouraging.

Optimal Instruments (see Conlon Gortmaker 2020)

BLP 1999 tells us the (Chamberlain 1987) optimal instruments for this supply-demand system of $G\Omega^{-1}$ where for a given observation n, we need to compute $E[\frac{\partial \xi_{jt}}{\partial \theta}|x, v, w]$ and $E[\frac{\partial \omega_{jt}}{\partial \theta}|x, v, w]$

$$D_{jt} \equiv \underbrace{\begin{bmatrix} \frac{\partial \xi_{jt}}{\partial \beta} & \frac{\partial \omega_{jt}}{\partial \beta} \\ \frac{\partial \xi_{jt}}{\partial \alpha} & \frac{\partial \omega_{jt}}{\partial \alpha} \\ \frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} & \frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2} \\ \frac{\partial \xi_{jt}}{\partial \gamma} & \frac{\partial \omega_{jt}}{\partial \gamma} \end{bmatrix}}_{(K_1+K_2+K_3)\times 2} = \begin{bmatrix} -x_{jt} & 0 \\ -v_{jt} & 0 \\ \frac{\partial \xi_{jt}}{\partial \alpha} & \frac{\partial \omega_{jt}}{\partial \alpha} \\ \frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} & \frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2} \\ 0 & -x_{jt} \\ 0 & -w_{jt} \end{bmatrix}, \quad \Omega_t \equiv \underbrace{\begin{bmatrix} \sigma^2_{\xi_t} & \sigma_{\xi_t \omega_t} \\ \sigma_{\xi_t \omega_t} & \sigma^2_{\omega_t} \end{bmatrix}}_{2\times 2}.$$

Optimal Instruments: (see Conlon Gortmaker 2020)

I replace co-linear elements with zeros using $\odot\Theta$

$$(D_{jt}\Omega_t^{-1}) \odot \Theta = \frac{1}{\sigma_\xi^2\sigma_\omega^2 - \sigma_{\xi\omega}^2} \cdot \begin{bmatrix} -\sigma_\omega^2 x_{jt} & 0 \\ -\sigma_\omega^2 v_{jt} & \sigma_{\xi\omega} v_{jt} \\ \sigma_\omega^2 \frac{\partial \xi_{jt}}{\partial \alpha} - \sigma_{\xi\omega}\frac{\partial \omega_{jt}}{\partial \alpha} & \sigma_\xi^2 \frac{\partial \omega_{jt}}{\partial \alpha} - \sigma_{\xi\omega}\frac{\partial \xi_{jt}}{\partial \alpha} \\ \sigma_\omega^2 \frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} - \sigma_{\xi\omega}\frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2} & \sigma_\xi^2 \frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2} - \sigma_{\xi\omega}\frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} \\ 0 & -\sigma_\xi^2 x_{jt} \\ \sigma_{\xi\omega} w_{jt} & -\sigma_\xi^2 w_{jt} \end{bmatrix}.$$

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

$$Z_{jt}^{\mathsf{Opt},D} \equiv \underbrace{E[(D_{jt}(Z_t)\Omega_t^{-1} \odot \Theta)_{\cdot 1}|Z_t]}_{K_1+K_2+(K_3-K_x)}, \quad Z_{jt}^{\mathsf{Opt},S} \equiv \underbrace{E[(D_{jt}(Z_t)\Omega_t^{-1} \odot \Theta)_{\cdot 2}|Z_t]}_{K_2+K_3+(K_1-K_x)}.$$

Aside: What does Supply tell us about Demand?

Demand

$$\begin{aligned}\partial\alpha :& \sigma_\omega^2 \frac{\partial \xi_{jt}}{\partial \alpha} - \sigma_{\xi\omega}\frac{\partial \omega_{jt}}{\partial \alpha}\\ \partial\sigma :& \sigma_\omega^2 \frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} - \sigma_{\xi\omega}\frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2}\end{aligned}$$

Supply

$$\begin{aligned}&\sigma_\xi^2 \frac{\partial \omega_{jt}}{\partial \alpha} - \sigma_{\xi\omega}\frac{\partial \xi_{jt}}{\partial \alpha}\\ &\sigma_\xi^2 \frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2} - \sigma_{\xi\omega}\frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2}\end{aligned}$$

- These are cross equation restrictions
- They serve as overidentifying restrictions for θ_2 parameters.
- This is the what imposing supply side tells us about demand (and *vice versa*)

Optimal Instruments

How to construct optimal instruments in form of Chamberlain (1987). Start with initial instruments $Z_{jt} = A\left(\mathbf{X_t}, \mathbf{W_t}, \mathbf{V_t}\right)$

$$E\left[\frac{\partial \xi_{jt}}{\partial \theta} | Z_{jt}\right] = \left[\beta, E\left[\frac{\partial \xi_{jt}}{\partial \alpha} | Z_{jt}\right], E\left[\frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} | Z_{jt}\right]\right]$$

Some challenges:

1. p_{jt} or η_{jt} depends on (ω_j, ξ_t) in a highly nonlinear way (no explicit solution!).
2. $E\left[\frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} \mid X_t, w_t\right] = E\left[\left[\frac{\partial \mathbf{s_t}}{\partial \delta_\mathbf{t}}\right]^{-1}\left[\frac{\partial \mathbf{s_t}}{\partial \widetilde{\theta}_2}\right] \mid Z_{jt}^D\right]$ (not conditioned on endogenous p!)

Feasible Recipe (BLP 1999)

1. Fix $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ and draw $(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*)$ from empirical density
2. Solve firm FOC's for $\hat{\mathbf{p}}_{\mathbf{t}}(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \widehat{\theta})$
3. Solve shares $\mathbf{s}_{\mathbf{t}}(\hat{\mathbf{p}}_{\mathbf{t}}, \widehat{\theta})$
4. Compute necessary Jacobian
5. Average over multiple values of$(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*)$. (Lazy approach: use only $(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*) = 0$).

In simulation the "lazy" approach does just as well.
(Caveat: At least for iid normally distributed $(\boldsymbol{\xi}, \boldsymbol{\omega})$)

Simplified Version: Reynaert Verboven (2014)

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

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

- Linear cost function means linear reduced-form price function (could do nonlinear regression too)

$$\begin{aligned} E\left[\frac{\partial \xi_{jt}}{\partial \alpha} | z_t\right] &= E[p_{jt}|z_t] = x_{jt}\gamma_1 + w_{jt}\gamma_2 \\ E\left[\frac{\partial \omega_{jt}}{\partial \alpha} | z_t\right] &= 0, \quad E\left[\frac{\partial \omega_{jt}}{\partial \widetilde{\theta}_2} | z_t\right] = 0 \\ E\left[\frac{\partial \xi_{jt}}{\partial \widetilde{\theta}_2} | z_t\right] &= E\left[\frac{\partial \delta_{jt}}{\partial \widetilde{\theta}_2} | z_t\right] \end{aligned}$$

- If we are worried about endogenous oligopoly markups is this a reasonable idea?
- Turns out that the important piece tends to be shape of jacobian for σ_x.

Optimal Instruments: Reynaert Verboven (2014)

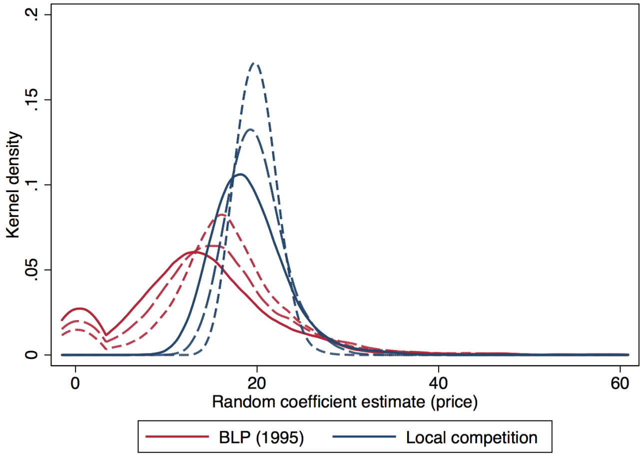
Table 2: Bias and Efficiency with Imperfect Competition

		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)

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



IV Comparison: Conlon and Gortmaker (2019)

				True Value				Median Bias				Median Absolute Error			
Simulation	Supply	Instruments	Seconds	α	σ_x	σ_p	ρ	α	σ_x	σ_p	ρ	α	σ_x	σ_p	ρ
Simple	No	Own	0.6	-1	3			0.126	-0.045			0.238	0.257		
Simple	No	Sums	0.6	-1	3			0.224	-0.076			0.257	0.208		
Simple	No	Local	0.6	-1	3			0.181	-0.056			0.242	0.235		
Simple	No	Quadratic	0.6	-1	3			0.206	-0.085			0.263	0.239		
Simple	No	Optimal	0.8	-1	3			0.218	-0.049			0.250	0.174		
Simple	Yes	Own	1.4	-1	3			0.021	0.006			0.226	0.250		
Simple	Yes	Sums	1.5	-1	3			0.054	-0.020			0.193	0.196		
Simple	Yes	Local	1.4	-1	3			0.035	-0.006			0.207	0.229		
Simple	Yes	Quadratic	1.4	-1	3			0.047	-0.022			0.217	0.237		
Simple	Yes	Optimal	2.2	-1	3			0.005	0.012			0.170	0.171		
Complex	No	Own	1.1	-1	3	0.2		-0.025	0.000	-0.200		0.381	0.272	0.200	
Complex	No	Sums	1.1	-1	3	0.2		0.225	-0.132	-0.057		0.263	0.217	0.200	
Complex	No	Local	1.0	-1	3	0.2		0.184	-0.107	-0.085		0.274	0.236	0.200	
Complex	No	Quadratic	1.0	-1	3	0.2		0.200	-0.117	-0.198		0.299	0.243	0.200	
Complex	No	Optimal	1.6	-1	3	0.2		0.191	-0.119	0.001		0.274	0.195	0.200	
Complex	Yes	Own	3.9	-1	3	0.2		-0.213	0.060	0.208		0.325	0.263	0.208	
Complex	Yes	Sums	3.3	-1	3	0.2		0.018	-0.104	0.052		0.203	0.207	0.180	
Complex	Yes	Local	3.4	-1	3	0.2		-0.043	-0.078	0.135		0.216	0.225	0.200	
Complex	Yes	Quadratic	3.5	-1	3	0.2		-0.028	-0.067	0.116		0.237	0.227	0.200	
Complex	Yes	Optimal	4.9	-1	3	0.2		-0.024	-0.036	-0.002		0.193	0.171	0.191	

IV Comparison: Conlon and Gortmaker (2019)