Aggregation and Endogeneity (BLP)

Chris Conlon

Fall 2029

NYU Stern

First: Some Review

What is the goal?

Consider the multi-product Bertrand problem where firms solve: $\arg\max_{p\in\mathcal{J}_f} \pi_f(\mathbf{p}) = \sum_{j\in\mathcal{J}_f}(p_j - c_j)\cdot q_j(\mathbf{p})$:

$$0 = q_j(\mathbf{p}) + \sum_{k\in\mathcal{J}_f}(p_k - c_k)\frac{\partial q_k}{\partial p_j}(\mathbf{p})$$

$$\rightarrow p_j = q_j(\mathbf{p})\left[-\frac{\partial q_j}{\partial p_j}(\mathbf{p})\right]^{-1} + c_j + \sum_{k\in\mathcal{J}_f\setminus j}(p_k - c_k)\underbrace{\frac{\partial q_k}{\partial p_j}(\mathbf{p})\left[-\frac{\partial q_j}{\partial p_j}(\mathbf{p})\right]^{-1}}_{D_{jk}(\mathbf{p})}$$

$$p_j(p_{-j}) = \underbrace{\frac{1}{1+1/\epsilon_{jj}(\mathbf{p})}}_{\text{Markup}}\left[c_j + \sum_{k\in\mathcal{J}_f\setminus j}(p_k - c_k)\cdot D_{jk}(\mathbf{p})\right].$$

We call $D_{jk}(\mathbf{p}) = \frac{\frac{\partial q_k}{\partial p_j}(\mathbf{p})}{\left|\frac{\partial q_j}{\partial p_j}(\mathbf{p})\right|}$ the diversion ratio and ϵ_{jj} the own elasticity and these are the main deliverables.

Starting Point: McFadden and MLE

Each individual's choice $d_{ij} \in \{0,1\}$ and $\sum_{j \in \mathcal{J}} d_{ij} = 1$.

Consumers make mutually exclusive and exhaustive choices to maximize (indirect) utility:

$$u_{ij} = \beta_i \, x_{ij} + \varepsilon_{ij} \text{ and } u_{i0} = \varepsilon_{i0}$$

$$d_{ij} = 1 \text{ IFF } [u_{ij} > u_{ik} \, \forall k \neq j]$$

Choices follow a Categorical distribution:

$$(d_{i1}, \dots, d_{iJ}, d_{i0}) \sim \text{Categorical}(s_{i1}, \dots, s_{iJ}, s_{i0})$$

Starting Point: McFadden and MLE

If we assume that ε_{ij} is Type I extreme value and $\beta_\iota \sim f(\beta_\iota \mid \theta)$ (some known parametric distribution) then we can write:

$$s_{ij}(\theta) = \mathbb{P}(d_{ij} = 1) = \int \frac{\exp[\beta_\iota x_j]}{1 + \sum_{k \in \mathcal{J}} \exp[\beta_\iota x_k]} f(\beta_\iota \mid \theta) \, \partial \beta_\iota$$

Which gives us the log-likelihood:

$$\ell(\theta) = \sum_{i=1}^{N} \sum_{j \in \mathcal{J} \cup \{0\}} d_{ij} \log s_{ij}(\mathbf{x_i} \mid \theta)$$

There are a bunch of challenges, not least among which is that the above is inconsistent if the integral is evaluated with error that doesn't decrease in N.

If each individual is exchangable then ex-ante they have the same choice probabilities: $s_{ij} = s_j$, and the sum of M Categoricals is Multinomial:

$$(q_1^*, \ldots, q_J^*, q_0^*) \sim \text{Mult}(M, s_1, \ldots, s_J, s_0)$$

where $q_j^* = \sum_{i=1}^{M} d_{ij}$ is a sufficient statistic.

- If M gets large enough then $\left(\frac{q_1}{M}, \ldots, \frac{q_J}{M}, \frac{q_0}{M}\right) \rightarrow (\mathfrak{s}_1, \ldots, \mathfrak{s}_J, \mathfrak{s}_0)$
- Idea: Equate observed market shares to the conditional choice probabilities $(s_1(\mathbf{x_i}, \theta), \ldots, s_J(\mathbf{x_i}, \theta), s_0(\mathbf{x_i}, \theta))$.
- Challenges: We probably don't really observe q_0 and hence M.
- Introduce idea of market t (otherwise not much data!)

Lots of papers stop here

Table IV. Hospital demand results, ML estimation

Interaction Terms	Variable	Estimated coefficient
	Distance (miles)	−0.215** (0.004)
	Distance squared	0.001** (0.000)
	Emergency* distance	−0.008** (0.004)
Interactions: Teaching	Cardiac	0.090 (0.060)
	Cancer	0.192** (0.069)
	Neurological	0.546** (0.175)
	Digestive	−0.145** (0.062)
	Labor	0.157** (0.048)
	Newborn baby	0.038 (0.075)
	Income ($000)	0.007** (0.001)
	PPO enrollee	−0.067 (0.050)
Interactions: Nurses per bed	Cardiac	−0.096 (0.070)
	Cancer	0.445** (0.079)
	Neurological	0.130 (0.200)
	Digestive	−0.028 (0.076)
	Labor	−0.002 (0.063)
	Newborn baby	0.071 (0.087)
	Income ($000)	0.005** (0.001)
	PPO enrollee	−0.099* (0.056)
Interactions: For-profit	Cardiac	−0.164 (0.181)
	Cancer	−0.197 (0.202)
	Neurological	0.229 (0.379)
	Digestive	0.195 (0.150)
	Labor	0.300** (0.107)
	Newborn baby	0.194* (0.122)
	Income ($000)	−0.001 (0.003)
	PPO enrollee	−0.036 (0.090)
Interactions: Cardiac services	Cardiac	1.222** (0.134)
	Income ($000)	0.001 (0.001)
	PPO enrollee	0.080 (0.088)
Interactions: Imaging services	Cardiac	−0.188** (0.094)
	Cancer	−0.052 (0.107)
	Neurological	−0.084 (0.287)
	Digestive	−0.182* (0.105)
	Labor	−0.071 (0.084)
	Newborn baby	0.398** (0.129)
	Income ($000)	0.004** (0.001)
	PPO enrollee	−0.061 (0.072)
Interactions: Cancer services	Cancer	0.073 (0.082)
	Income ($000)	−0.005** (0.001)
	PPO enrollee	0.087 (0.056)
Interactions: Labor services	Labor	3.544** (0.391)
	Newborn baby	3.116** (0.487)
	Income ($000)	−0.003* (0.002)
	PPO enrollee	0.045 (0.077)
	Hospital fixed effects	Yes
	Pseudo-R^2	0.43

Notes: Maximum likelihood estimation of demand for hospitals using a multinomial logit model. Specification includes hospital fixed effects. $N = 28\,666$ encounters. Standard errors are reported in parentheses. ** Significant at $p = 0.05$; * Significant at $p = 0.1$.

- This just MLE on the full individual data from Ho (2006)
  - There is no unobserved heterogeneity, just a deterministic $\beta(y_i)$ where y_i are demographics.
- The “FTC model” (Raval, Rosenbaum, Wilson RJE 2022)/ (Raval, Rosenbaum, Tenn EI 2017) groups individuals by income, diagnosis, and zip code and estimates a separate set of $\beta_{g(i)}$ for each group.
- Hospitals are a bit special: distance x_{ij} does much of the work (special regressor)
- Price endogeneity not really a concern (?)

$$\min_{\pi_i} \sum_{j,t} \left(\mathcal{S}_{jt}(\mathbf{x_t}) - \sum_i \pi_i \cdot s^*_{ijt}(\beta^*_i, \mathbf{x_t}) \right)^2 \quad \text{subject to} \quad s^*_{ijt}(\beta^*_i, \mathbf{x_t}) = \frac{e^{\beta^*_i x_{jt}}}{1 + \sum_{j'} e^{\beta^*_i x_{j't}}}$$

$$0 \leq \pi_i \leq 1, \quad \sum_i \pi_i = 1$$

1. Draw a large number (thousands?) of $(\beta^*_i$ from a prior distribution $g(\beta_i)$ more dispersed than the true $f(\beta_i)$
2. Compute individual choice probabilities $s^*_{ijt}(\beta^*_i)$
3. Estimate above by constrained least squares (non-negative lasso)
4. This produces sparse models. (most $\pi_i = 0$)

Fixed coefficients require EM. See Heiss, Hetzenecker, Osterhaus (JE 2022) for details (and elastic net variant $\sum_i \pi_i^2 \leq t$).

Applications of Fixed Grid Estimators

These estimators are helpful when computing choice probabilities is time-consuming (ie: when choices are dynamic). Some recent examples:

- Nevo, Turner, Williams (ECMA 2016): Broadband competition
- Blundell, Gowrisankaran, Langer (AER 2020): EPA regulation

Review: “Classic” BLP (1995) Models

Inversion: IIA Logit

Add unobservable error for each $\mathfrak{s}_{jt}$ labeled ξ_{jt}.

$$u_{ijt} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}} + \varepsilon_{ijt}, \quad \sigma_j(\boldsymbol{\delta_t}) = \frac{e^{\delta_{jt}}}{1 + \sum_k e^{\delta_{kt}}}$$

- The idea is that ξ_{jt} is observed to the firm when prices are set, but not to us the econometricians.
- Potentially correlated with price Corr$(\xi_{jt}, p_{jt}) \neq 0$
- But not characteristics $\mathbb{E}[\xi_{jt} \mid x_{jt}] = 0$.
  - 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.
  - Something about a BMW makes it better than a Peugeot but is not fully captured by characteristics that leads higher sales and/or higher prices.
  - Consumers agree on its value (vertical component).

Inversion: IIA Logit

Taking logs:

$$\ln s_{0t} = -\log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right)$$

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

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

Exploit the fact that:

1. $\ln s_{jt} - \ln s_{0t} = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t}$ (with no sampling error)
2. We have one ξ_{jt} for every share s_{jt} (one to one mapping)

Inversion: Nested Logit (Berry 1994 / Cardell 1991)

This takes a bit more algebra but not much

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

$$\ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} = x_{jt}\beta - \alpha p_{jt} + \rho \log(\mathfrak{s}_{j|gt}) + \xi_{jt}$$

- Same as logit plus an extra term $\log(s_{j|g})$ the within group share.
  - We now have a second endogenous regressor.
  - 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 ρ you will get $\rho \to 1$ because of attenuation bias.
- A common instrument for ρ is the number of products within the nest. Why?

Think about a generalized inverse for $\sigma_j(\boldsymbol{\delta_t}, \mathbf{x}_t, \theta_2) = \mathfrak{s}_{jt}$ so that

$$\sigma_{jt}^{-1}(\mathcal{S}_{\cdot t}, \tilde{\theta}_2) = \delta_{jt} \equiv x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

- After some transformation of data (shares $\mathcal{S}_{\cdot t}$) we get mean utilities δ_{jt}.
- Same IV-GMM approach after transformation
- Examples:
  - Plain Logit: $\sigma_j^{-1}(\mathcal{S}_{\cdot t}) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t}$
  - Nested Logit: $\sigma_j^{-1}(\mathcal{S}_{\cdot t}, \rho) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} + \rho \ln \mathfrak{s}_{j|gt}$
  - Three level nested logit: $\sigma_j^{-1}(\mathcal{S}_{\cdot t}, \rho) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} + \sum_{d=1}^{2} \rho_d \ln\left(\frac{s_{jt}}{s_{d(j),t}}\right)$ (Verboven 1996)
- Anything with a share requires an IV (otherwise $\rho \to 1$).

Fosgerau, Monardo, De Palma (2022) propose the IPDL

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}\beta - \alpha p_{jt} + \sum_{d=1}^{D} \rho_d \ln\left(\frac{s_{jt}}{s_{d(j),t}}\right) + \xi_{jt}$$

- Can accomodate multiple (partially) overlapping nests
- The naive idea (include share within group on RHS with IV) actually works like you would want (!)
- We need that $\rho > 0$ for this to be RUM.
- This can allow for mild complementarities as well.

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

$$\sigma_j(\boldsymbol{\delta_t}; \mathbf{x_t}; \theta_2) = \int \frac{\exp[\delta_{jt} + \mu_{ij}]}{1 + \sum_k \exp[\delta_{kt} + \mu_{ik}]} f(\boldsymbol{\mu_i} | \theta_2)$$

- We typically parametrize $\mu_{ijt} = x_{jt} \cdot [\Pi\, y_i + \Sigma\, \nu_i]$ where y_i are demographics and ν_i are unobserved heterogeneity (typically multivariate normal).
- Label $\theta_2 = [\Pi, \Sigma, \alpha]$
- 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.

Lots of ways to solve equations (Conlon Gortmaker 2020)

- If you can work out $\frac{\partial \sigma_{jt}}{\partial \delta_{kt}}$ (easy) you can solve this using Newton's Method.
- BLP prove (not easy) that this is a contraction mapping.

$$\boldsymbol{\delta}^{(\mathbf{k})}(\theta) = \boldsymbol{\delta}^{(\mathbf{k-1})}(\theta) + \log(\mathcal{S}_j) - \log(\sigma_j(\boldsymbol{\delta}_{\mathbf{t}}^{(k-1)}, \theta))$$

  - Practical tip: ϵ_{tol} needs to be as small as possible. ($\approx 10^{-13}$).
  - Practical tip: Contraction isn't as easy as it looks: $\log(\sigma_j(\delta_t^{(k-1)}, \theta))$ requires computing the numerical integral each time (either via quadrature or monte carlo).

- We can use accelerated fixed point techniques (SQUAREM) (see Reynaerts, Varadhan, and Nash 2012). [PyBLP default].

Rewriting as a Convex Program (Li 2018, Monardo 2024)

We can also solve the following convex program:

$$\min_{\boldsymbol{\delta}} \sum_{i=1}^{I} w_i \cdot \log\left(\sum_j \exp\left(\delta_j + \mu_{ij}(\theta_2)\right)\right) - \sum_k \delta_k \cdot \mathfrak{s}_k$$

The first-order conditions are given by

$$\sigma_j(\boldsymbol{\delta}, \theta_2) = \mathfrak{s}_j$$

And the Hessian requires knowledge of $\frac{\partial \sigma_j}{\partial \delta_k}$

- We are back to solving the non-linear system of equations, but...
- Convex programming might be faster/more robust than root finding (why?)
- This makes the proof of a unique solution trivial...

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:
  - 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 $W \rightarrow W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

The model is still defined by CMR $\mathbb{E}[\xi_{jt} \mid z_{jt}^D] = 0$

- 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 $\hat{\alpha}(\theta_2), \hat{\beta}(\theta_2), \hat{\xi}(\theta_2)$.
- Outer Loop update guess θ, solve for δ and repeat.

$$\widehat{\theta_{BLP}} = \arg\max_{\theta}(Z'\hat{\xi}(\theta_2))W(Z'\hat{\xi}(\theta_2))'$$

- When we have found $\widehat{\theta_2}_{BLP}$ we can use this to update $W \to W(\widehat{\theta_2}_{BLP})$ and do 2-stage GMM.

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

$$\delta_{jt}(\theta_2) = x_{jt}\beta - \alpha p_{jt} + \underbrace{\xi_{jt}}_{\xi_j+\xi_t+\Delta\xi_{jt}}$$

- 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.
  - Good: endogeneity problem less severe.
  - Bad: less variation in IV.

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 $\delta(\theta_2), \xi(\theta_2)$ 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.

$$\begin{aligned} \arg\min_{\theta_2} \psi' \Omega^{-1} \psi \quad \text{s.t.} \\ \psi = \xi(\theta_2)' Z \\ \xi_{jt}(\theta_2) = \delta_{jt}(\theta_2) - x_{jt}\beta - \alpha p_{jt} \\ \log(\mathcal{S}_{jt}) = \log(\sigma_{jt}(\delta, \theta_2)) \end{aligned} \tag{1}$$

$$\begin{aligned} \arg\min_{\theta_2,\alpha,\beta,\xi,\psi} & \quad \psi' \Omega^{-1} \psi \quad \text{s.t.} \\ \psi &= \xi' Z \\ \xi_{jt} &= \delta_{jt} - x_{jt}\beta - \alpha p_{jt} \\ \log(\mathcal{S}_{jt}) &= \log(\sigma_{jt}(\theta_2, \delta)) \end{aligned} \tag{2}$$

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).
  - $\widehat{\theta}_{NFP} \approx \widehat{\theta}_{MPEC}$ but for numerical differences in the optimization routine.
- Our choice of algorithm should mostly be about computational convenience.

- Advantages
  - Concentrate out all of the linear in utility parameters (ξ, δ, β) so that we only search over θ_2. When dim$(\Sigma) = \theta_2$ is small (few dimensions of unobserved heterogeneity) this is a big advantage. For $K \leq 5$ 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.

BLP: MPEC Advantages/Disadvantages

- Advantages
  - Problem scales better in dim(θ_2).
  - Because all constraints hold at the optimum only: less impact of numerical error in tolerance or integration.
  - Derivatives are less complicated than $\frac{\partial \delta}{\partial \theta_2}$ (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 $\rightarrow$ Fixed effects are harder
  - We have to find the derivatives of the shares with respect to all of the parameters β, ξ, θ_2. (The other derivatives are pretty easy).
  - Parallelizing the derivatives is trickier than NFP case.

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).

Some setup

We can break up the parameter space into three parts:

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

Supply Side

Consider the multi-product Bertrand FOCs:

$$\arg\max_{p\in\mathcal{J}_f} \pi_f(\mathbf{p}) = \sum_{j\in\mathcal{J}_f}(p_j - c_j)\cdot s_j(\mathbf{p}) + \kappa_{fg}\sum_{k\in\mathcal{J}_g}(p_k - c_k)\cdot s_k(\mathbf{p})$$

$$0 = s_j(\mathbf{p}) + \sum_{k\in\mathcal{J}_f}(p_k - c_k)\frac{\partial s_k}{\partial p_j}(\mathbf{p})$$

It is helpful to define the ownership matrix:

$$\mathcal{H}(\kappa)_{(j,k)} = \left\{\begin{array}{lr} 1 & \text{for } (j,k)\in\mathcal{J}_f \text{ for any } f \\ 0 & \text{o.w} \end{array}\right\}$$

We can re-write the FOC in matrix form where $\odot$ denotes Hadamard product (element-wise) and $\Delta_{(j,k)}(\mathbf{p}) = -\frac{\partial s_j}{\partial p_k}(\mathbf{p})$ is the demand derivatives matrix:

$$s(\mathbf{p}) = (\mathcal{H}\odot\Delta(\mathbf{p}))\cdot(\mathbf{p} - \mathbf{mc}),$$

$$\mathbf{mc} = \mathbf{p} - \underbrace{(\mathcal{H}\odot\Delta(\mathbf{p}))^{-1}s(\mathbf{p})}_{\eta(\mathbf{p},\mathbf{s},\theta_2)}.$$

Recovering Marginal Costs

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

$$\mathbf{mc}(\theta) = \mathbf{p} - \boldsymbol{\eta}(\mathbf{p}, \mathbf{s}, \theta_2)$$
$$f(mc_{jt}) = [\mathsf{x}_{jt}, \mathsf{w}_{jt}]\theta_3 + \omega_{jt}$$

Which we can solve for ω_{jt}:

$$\omega_{jt} = f(\mathbf{p} - \boldsymbol{\eta}(\mathbf{p}, \mathbf{s}, \theta_2)) - [\mathsf{x}_{jt}, \mathsf{w}_{jt}]\theta_3$$

- $f(\cdot)$ is usually $\log(\cdot)$ or identity.
- We can use this to form additional moments: $\mathbb{E}[\omega_{jt} \mid z_{jt}^s] = 0$.
- We can just stack these up with the demand moments $E[\xi'_{jt} Z_{jt}^d] = 0$.
- This step is optional but can aid in identification (if you believe it).

Some different definitions:

$$
\begin{aligned}
y_{jt}^{D} &:= \hat{\delta}_{jt}(\theta_2) + \alpha p_{jt} &&= (\mathsf{x}_{jt}\ \mathsf{v}_{jt})'\beta + \xi_t =: x_{jt}^{D\prime}\beta + \xi_{jt} \\
y_{jt}^{S} &:= \widehat{mc}_{jt}(\theta_2) &&= (\mathsf{x}_{jt}\ \mathsf{w}_{jt})'\gamma + \omega_t =: x_{jt}^{S\prime}\gamma + \omega_{jt}
\end{aligned} \tag{3}
$$

Stacking the system across observations yields:

$$
\underbrace{\begin{bmatrix} y_D \\ y_S \end{bmatrix}}_{2N\times 1} = \underbrace{\begin{bmatrix} X_D & 0 \\ 0 & X_S \end{bmatrix}}_{2N\times(K_1+K_3)} \underbrace{\begin{bmatrix} \beta \\ \gamma \end{bmatrix}}_{(K_1+K_3)\times 1} + \underbrace{\begin{bmatrix} \xi \\ \omega \end{bmatrix}}_{2N\times 1} \tag{4}
$$

Simultaneous Supply and Demand: in details

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

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

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

(d) Stack up $\hat{\delta}_{\cdot t}(\theta_2)$ and $\widehat{mc}_{jt}(\hat{\delta}_{\cdot t}(\theta_2), \theta_2)$ and use linear IV-GMM to recover $[\hat{\theta}_1(\theta_2), \hat{\theta}_3(\theta_2)]$ following the recipe in Appendix of Conlon Gortmaker (2020)

(e) Construct the residuals:

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

(f) Construct sample moments

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

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

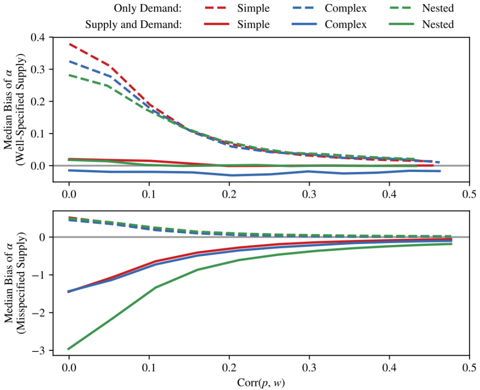
What's the Point (Conlon Gortmaker RJE 2020)

- A well-specified supply side can make it easier to estimate θ_2 parameters (price in particular).
- Imposing the supply side only helps if we have information about the marginal costs / production function that we would like to impose
- May want to enforce some economic constraints: ($mc_{jt} > 0$ is a good one).
- But assuming the wrong conduct $\mathscr{H}_t$ can lead to misspecification!

				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		

What about Misspecification? (Conlon Gortmaker RJE 2020)

Figure 2: Instrument Strength and Misspecification



Each plot documents how bias of the linear parameter on price, α, decreases with the strength of the cost shifter w_{jt}, which is included as a demand-side instrument. To weaken or strengthen the instrument, we vary its supply-side parameter from $\gamma_w = 0$ to $\gamma_w = 1$, and report the correlation this induces between w_{jt} and prices p_{jt}. Reported bias values are medians across 1,000 different simulations. The top plot reports results for the simulation configurations described in Section 5. In the bottom plot, we simulate data according to perfect competition (i.e., prices are set equal to marginal costs instead of those that satisfy Bertrand-Nash first order conditions), but continue to estimate the model under the assumption of imperfect competition. For all problems, we use the "approximate" version of the feasible optimal instruments and a Gauss-Hermite product rule that exactly integrates polynomials of degree 17 or less.

Quick Case Study

$$\mathbf{mc} = \mathbf{p} - \eta(\mathbf{p}, \mathbf{s}, \theta_2).$$

Demand systems have two main deliverables:

- Own-price elasticities $\epsilon_{jj}(\mathbf{p}, \theta_2)$
- Substitution patterns (all transforms of same object)
  - Demand Derivatives: $\Delta(\mathbf{p}, \theta_2)$
  - Cross elasticities: $\epsilon_{jk}(\mathbf{p}) = \frac{\partial q_k}{\partial p_j}$
  - Diversion Ratios: $D_{jk}(\mathbf{p}) = \frac{\partial q_k}{\partial p_j} / |\frac{\partial q_j}{\partial p_j}|$
- Other checks: $D_{j0}(\mathbf{p})$ diversion to outside good; ϵ^{agg} category elasticity to 1% tax.
  - Tend to give insight into the economics of the problem.

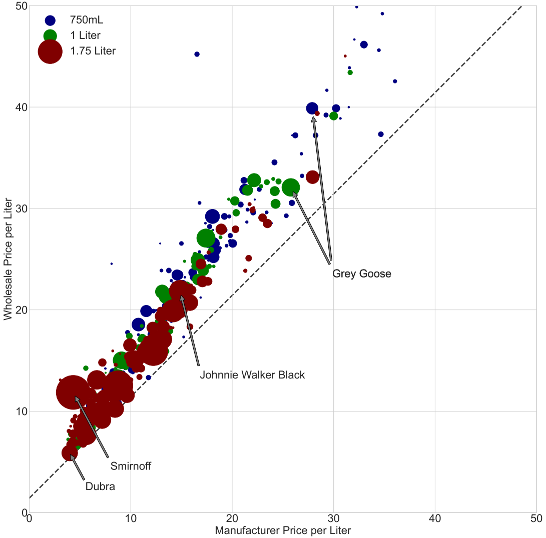
Consumer i chooses product j (brand-size-flavor) in quarter t:

$$u_{ijt} = \beta_i^0 - \alpha_i\, p_{jt} + \beta_i^{1750} \cdot \mathbb{I}[1750mL]_j + \gamma_j + \gamma_t + \varepsilon_{ijt}(\rho)$$

$$\begin{pmatrix} \ln \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \overline{\alpha} \\ \theta_1 \end{pmatrix} + \Sigma \cdot \nu_i + \sum_k \Pi_k \cdot \mathbb{I}\{LB_k \leq \text{Income}_i < UB_k\}$$

- Nesting Parameter ρ: Substitution within category (Vodka, Gin, etc.)
- Consumers of different income levels have different mean values for coefficients
- Conditional on income, normally distributed unobserved heterogeneity for:
  - Price α_i
  - Constant β_i^0 (Overall demand for spirits)
  - Package Size: β_i^{1750} (Large vs. small bottles)

Wholesale Margins Under Post and Hold



- Price Cost Margins (and Lerner Markups) are higher on premium products
- Markups on least expensive products (plastic bottle vodka) are very low.
- Smirnoff (1.75L) is best seller (high markup / outlier).
- A planner seeking to minimize ethanol consumption would flatten these markups!
- Matching this pattern is kind of the whole ballgame !
- Plain logit gives $\epsilon_{jj} = \alpha \cdot p_j \cdot (1 - s_j)$.

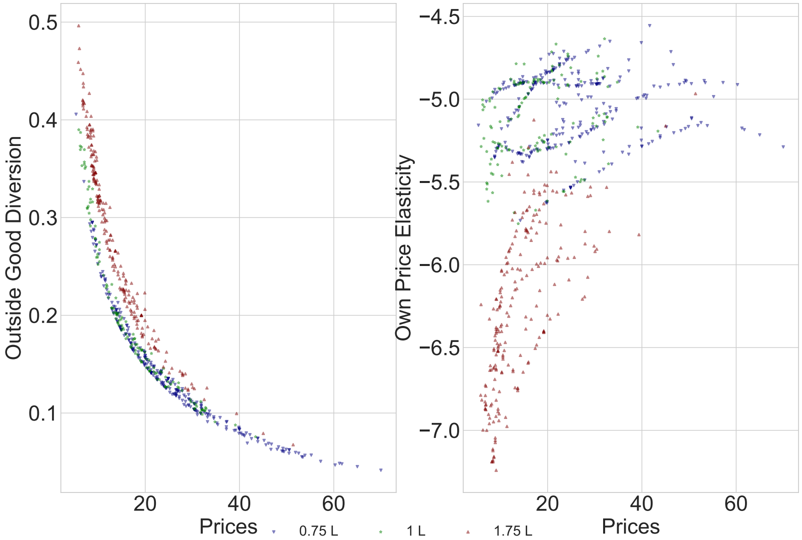
Demand Estimates (from PyBLP, Conlon Gortmaker (2020, 2023))

Π	Const	Price	1750mL
Below $25k	2.928	-0.260	0.543
	(0.233)	(0.056)	(0.075)
$25k-$45k	0.184	-0.170	0.536
	(0.236)	(0.054)	(0.083)
$45k-$70k	0.000	-0.179	0.980
	(0.000)	(0.053)	(0.093)
$70k-$100k	-0.452	-0.496	0.608
	(0.227)	(0.051)	(0.079)
Above $100k	-1.777	-1.543	0.145
	(0.234)	(0.047)	(0.055)
Σ^2			
Constant	1.167	0.695	
	(0.236)	(0.048)	
Price	0.695	0.697	
	(0.048)	(0.028)	
Nesting Parameter ρ		0.423	
		(0.026)	
Fixed Effects		Brand+Quarter	
Model Predictions	25%	50%	75%
Own Elasticity: $\frac{\partial \log q_j}{\partial \log p_j}$	-5.839	-5.162	-4.733
Aggregate Elasticity: $\frac{\partial \log Q}{\partial \log P}$	-0.333	-0.329	-0.322
Own Pass-Through: $\frac{\partial p_j}{\partial c_j}$	1.256	1.284	1.320
Observed Wholesale Markup (PH)	0.188	0.233	0.276
Predicted Wholesale Markup (PH)	0.205	0.231	0.259

- Demographic Interactions w/ 5 income bins (matched to micro-moments)
- Correlated Normal Tastes: (Constant, Large Size, Price)
- Supply moments exploit observed upstream prices and tax change (ie: match observed markups).

  $\mathbb{E}[\omega_{jt}] = 0,$ with $\omega_{jt} = \left(p^w_{jt} - p^m_{jt} - \tau_{jt}\right) - \eta_{jt}\left(\theta_2\right).$
- Match re-purchase rate of $Vodka \to Vodka$ for each category.
- Pass-through consistent with estimates from our AEJ:Policy paper.

Elasticities and Diversion Ratios



Diversion Ratios

	Median Price	% Substitution		Median Price	% Substitution
Capt Morgan Spiced 1.75 L ($15.85)			Cuervo Gold 1.75 L ($18.33)		
Bacardi Superior Lt Dry Rum 1.75 L	12.52	13.07	Don Julio Silver 1.75 L	22.81	5.00
Bacardi Dark Rum 1.75 L	12.52	2.71	Cuervo Gold 1.0 L	21.32	3.82
Bacardi Superior Lt Dry Rum 1.0 L	15.03	2.44	Sauza Giro Tequila Gold 1.0 L	8.83	3.07
Smirnoff 1.75 L	11.85	2.36	Smirnoff 1.75 L	11.85	2.44
Lady Bligh Spiced V Island Rum 1.75 L	9.43	2.18	Absolut Vodka 1.75 L	15.94	2.06
Woodford 0.75 L ($34.55)			Beefeater Gin 1.75 L ($17.09)		
Jack Daniel Black Label 1.0 L	27.08	7.66	Tanqueray 1.75 L	17.09	12.80
Jack Daniel Black Label 1.75 L	21.85	4.91	Gordons 1.75 L	11.19	4.14
Jack Daniel Black Label 0.75 L	29.21	4.83	Seagrams Gin 1.75 L	10.23	2.85
Makers Mark 1.0 L	32.79	4.52	Bombay 1.75 L	21.95	2.27
Makers Mark 0.75 L	31.88	2.80	Smirnoff 1.75 L	11.85	2.27
Dubra Vdk Dom 80P 1.75 L ($5.88)			Belvedere Vodka 0.75 L ($30.55)		
Popov Vodka 1.75 L	7.66	7.56	Grey Goose 1.0 L	32.08	5.09
Smirnoff 1.75 L	11.85	3.15	Absolut Vodka 1.75 L	15.94	3.82
Sobieski Poland 1.75 L	9.09	3.14	Absolut Vodka 1.0 L	24.91	2.74
Grays Peak Vdk Dom 1.75 L	9.16	2.87	Smirnoff 1.75 L	11.85	2.43
Wolfschmidt 1.75 L	6.92	2.48	Grey Goose 0.75 L	39.88	2.22

Instruments and Identification

- Once we have $\delta_{jt}(\theta)$ identification of linear parameters $\theta_1 = [\beta, \xi_j, \xi_t]$ 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.
- Intuition: How are θ_2 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 θ_2 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.

Instruments

- Recall the nested logit, where there are two separate endogeneity problems
  - Price: this is the familiar one!
  - Nonlinear characteristics/Shares θ_2 this is the other one.
- We are doing nonlinear GMM: Start with $\mathbb{E}[\xi_{jt}|x_{jt}, z_{jt}] = 0$ use $\mathbb{E}[\xi'[ZX]] = 0$.
  - In practice this means that for valid instruments (x, z) any function $f(x, z)$ is also a valid instrument $\mathbb{E}[\xi_{jt} f(x_{jt}, z_{jt})] = 0$.
  - 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?

Exclusion Restrictions (see Berry Haile 2014)

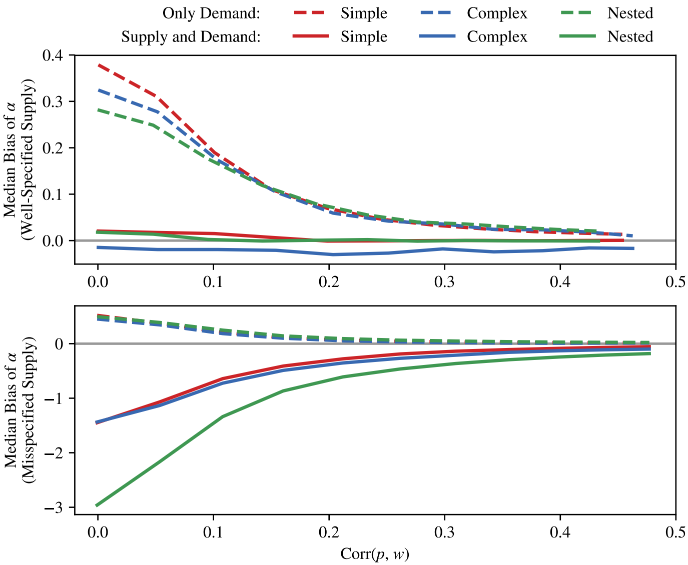
$$\begin{aligned}\delta_{jt}(\mathcal{S}_t, \mathsf{y}_t, \tilde{\theta}_2) &= [\mathsf{x}_{jt}, \mathsf{v}_{jt}]\beta - \alpha p_{jt} + \xi_{jt} \\ f(p_{jt} - \eta_{jt}(\theta_2, \mathbf{p}, \mathbf{s})) &= h(\mathsf{x}_{jt}, \mathsf{w}_{jt}; \theta_3) + \omega_{jt}\end{aligned}$$

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
- y_t is a sneaky demand shifter
- If it doesn't shift either is it really relevant?

Alternative: MacKay Miller (2022) propose $Cov(\xi_{jt}, \omega_{jt}) = 0$ as an alternative.

Cost Shifters Really Matter (from Conlon Gortmaker RJE)



What about Hausman Instruments?

AKA contemporaneous prices of same product in a different market.

- Idea is to pick up common cost shocks:

$$p_{jmt} = c_{jmt} + \eta_{jmt}$$

- But this places strong assumptions on nature of demand shocks (and markups η_{jmt})
- Even with FE: $\xi_{jmt} = \xi_j + \xi_t + \underbrace{\Delta\xi_{jt}}_{=0} + \Delta\xi_{jmt}$
- A common complaint: national advertising might increase demand for a product in multiple geographic markets.

Markup Shifters

The equilibrium markup is a function of everything! $\eta_{jt}(\mathbf{p}, \mathbf{s}, \xi_t, \omega_t, \mathsf{x}_t, \mathsf{w}_t, \mathsf{v}_t, \mathsf{y}_t, \theta_2)$:

- It is obviously endogenous (depends on error terms)!
- But lots of potential instruments beyond excluded v_t or w_t.
- Idea: cross-market variation in number or strength of competitors
  - Also v_{-j} and w_{-j} and x_{-j}.
  - Not p_{-j} or ξ_{-j}, etc.
  - The idea is that these instruments shift the marginal revenue curve.
  - What is a good choice of $f(x_{-j})$? etc.

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.

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.

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

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

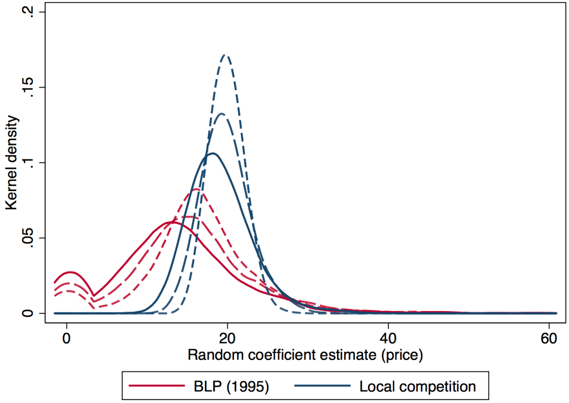
- And use this transformed to construct two kinds of IV (Squared distance, and count of local competitors)

$$DIV_1 = \sum_{j\in F} d_{jt}^2, \quad \sum_{j\notin F} d_{jt}^2$$
$$DIV_2 = \sum_{j\in F} \mathbb{I}[d_{jt} < c] \quad \sum_{j\notin F} \mathbb{I}[d_{jt} < c]$$

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

Differentiation Instruments: Gandhi Houde (2019)

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



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

Intuition from Linear IV (FRAC: Salanie and Wolak)

Simple case where $\theta_0 = (\beta_0, \pi_0, \sigma_0)'$. A second-order Taylor expansion around $\pi_0 = \sigma_0 = 0$ gives the following linear model with four regressors:

$$\log \frac{\mathcal{S}_{jt}}{\mathcal{S}_{0t}} \approx \beta_0 x_{jt} + \sigma_0^2 a_{jt} + \pi_0 m_t^y x_{jt} + \pi_0^2 v_t^y a_{jt} + \xi_{jt}, \quad a_{jt} = \left(\frac{x_{jt}}{2} - \sum_{k \in \mathcal{J}_t} \mathcal{S}_{kt} \cdot x_{kt} \right) \cdot x_{jt} \tag{5}$$

- $m_t^y = \sum_{i \in \mathcal{I}_t} w_{it} \cdot y_{it}$ is the within-market demographic mean
- $v_t^y = \sum_{i \in \mathcal{I}_t} w_{it} \cdot (y_{it} - m_t^y)^2$ is its variance
- a_{jt} is an "artificial regressor" that reflects within-market differentiation of the product characteristic x_{jt}.
- Linear but we still need an IV for a_{jt}.

Implemented in Julia by Jimbo Brand https://github.com/jamesbrandecon/FRAC.jl

Connection or when do GH IV work well?

Recall the GH IV are:

$$J \cdot x_{jt}^2 + \underbrace{\sum_k x_{kt}^2}_{\text{constant for } t} - 2 \sum_k x_{jt} \cdot x_{kt}$$

and the artificial regressor is

$$\frac{1}{2} x_{jt}^2 - 2 x_{jt} \cdot \sum_k \mathcal{S}_{kt} \cdot x_{kt}$$

- We should be share weighting the interaction term, but GH assume equal weighting.
- Should be able to do better than these IV (but ideal is infeasible...)
- Alternative take: GH propose IIA test that looks a lot like Salanie Wolak estimator. Good for starting values? Or as pre-test for heterogeneity?
- Warning: I find these are always nearly colinear and run PCA first...

- Since any $f(x,z)$ satisfies our orthogonality condition, we can try to choose $f(x,z)$ as a basis to approximate optimal instruments. (Newey 1990)
- 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 (Chamberlain 1987)

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

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

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

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:

$$G = \mathbb{E}\left[\frac{\partial g(z_i,\theta)}{\partial \theta}|z_i,\theta_0\right]$$

Optimal Instruments (Chamberlain 1987)

Consider the simplest IV problem:

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

Which gives:

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

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

Optimal Instruments (Newey 1990)

From previous slide, nothing says that $\mathbb{E}[x_i \mid v_i, z_i]$ 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. $(K \leq 5)$.
  - But recent improvements in sieves, LASSO, non-parametric regression are encouraging.

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

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

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

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 $\mathbb{E}[\frac{\partial \xi_{jt}}{\partial \theta}|\mathbf{Z_t}]$ and $\mathbb{E}[\frac{\partial \omega_{jt}}{\partial \theta}|\mathbf{Z_t}]$

$$G_{jt} \equiv \underbrace{\begin{bmatrix} \frac{\partial \xi_{jt}}{\partial \beta} & \frac{\partial \omega_{jt}}{\partial \beta} \\ \frac{\partial \xi_{jt}}{\partial \alpha} & \frac{\partial \omega_{jt}}{\partial \alpha} \\ \frac{\partial \xi_{jt}}{\partial \tilde{\theta}_2} & \frac{\partial \omega_{jt}}{\partial \tilde{\theta}_2} \\ \frac{\partial \xi_{jt}}{\partial \gamma} & \frac{\partial \omega_{jt}}{\partial \gamma} \end{bmatrix}}_{(K_1+K_2+K_3)\times 2} = \begin{bmatrix} -\mathsf{x}_{jt} & 0 \\ -\mathsf{v}_{jt} & 0 \\ \frac{\partial \xi_{jt}}{\partial \alpha} & \frac{\partial \omega_{jt}}{\partial \alpha} \\ \frac{\partial \xi_{jt}}{\partial \tilde{\theta}_2} & \frac{\partial \omega_{jt}}{\partial \tilde{\theta}_2} \\ 0 & -\mathsf{x}_{jt} \\ 0 & -\mathsf{w}_{jt} \end{bmatrix}, \quad \Omega_t \equiv \underbrace{\begin{bmatrix} \sigma^2_{\xi_t} & \sigma_{\xi_t \omega_t} \\ \sigma_{\xi_t \omega_t} & \sigma^2_{\omega_t} \end{bmatrix}}_{2\times 2}.$$

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

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

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

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

Optimal Instruments

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

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

Some challenges:

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

Things are infeasible because we don't know θ_0!

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 $\widehat{\mathbf{p}}_{\mathbf{t}}(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \widehat{\theta})$
3. Solve shares $\mathbf{s}_{\mathbf{t}}(\widehat{\mathbf{p}}_{\mathbf{t}}, \widehat{\theta})$
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. (At least for iid normal $(\boldsymbol{\xi}, \boldsymbol{\omega})$)

Alternative: Can we use $\mathbb{E}[\mathbf{p}_{\mathbf{t}} \mid \mathbf{Z}_{\mathbf{t}}]$ instead for (2) if we don't have supply side

Simplified Version: 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 (could do nonlinear regression too)

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

- 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.
- In either case what we care about is $\mathbb{E}[p \mid x, z]$ (the first stage). Nothing is free here !

Optimal Instruments: Reynaert Verboven (2014)

Table 2: Bias and Efficiency with Imperfect Competition

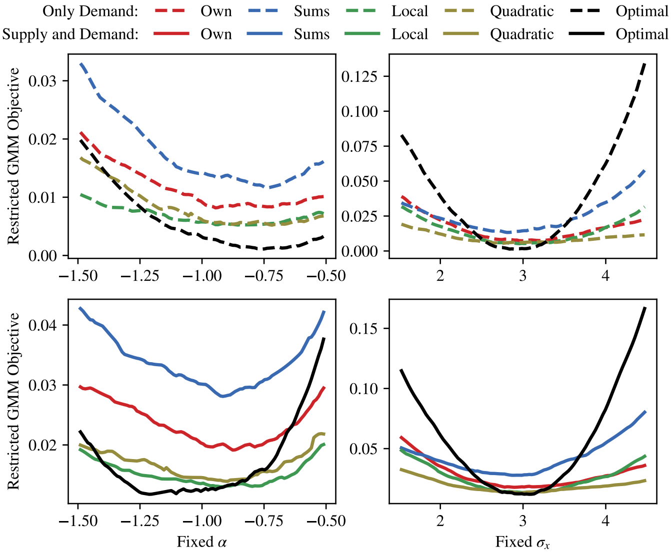
		Single Equation GMM								
			g_{jt}^1			g_{jt}^2			g_{jt}^3	
	True	Bias	St Err	RMSE	Bias	St Err	RMSE	Bias	St Err	RMSE
β^0	2	-0.127	0.899	0.907	-0.155	0.799	0.814	-0.070	0.514	0.519
β^1	2	-0.068	0.899	0.901	0.089	0.766	0.770	-0.001	0.398	0.398
α	-2	0.006	0.052	0.052	0.010	0.049	0.050	0.010	0.043	0.044
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.

IV Comparison: Conlon and Gortmaker (2020)

				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 (2020)



Takeaway

What does this mean:

- We should always check $\mathbb{E}[p \mid x, z]$ before we do anything else.
- Can use FRAC to figure out where the heterogeneity is, get startingn values
- May want to consider adding a supply side (if you're willing to assume for counterfactuals, why not?)
- Certainly should do `results.compute_optimal_instruments()` in PyBLP.