Horizontal Merger Simulation and Multiproduct Demand Estimation

Chris Conlon

Summer 2023

NYU Stern and NBER

Who am I?

I am an IO economist at NYU Stern School of Business, some work of mine:

- Does common ownership manifest in higher prices for consumer packaged goods? How much as common ownership grown in the US since 1980? What is the driver? (with Matt Backus and Mike Sinkinson)
- Can dominant firms use loyalty rebate contracts to exclude more efficient rivals? (with Julie Mortimer)
- Why is market power a bad way to address externalities? (with Nirupama Rao)
- What are most effective ways to estimate multi-product demand systems? Including with micro data (with Jeff Gortmaker)
- Do do different interventions (price changes, quality reduction, second-choices) lead us to estimate different diversion ratios? (with Julie Mortimer)
- Can we flexibly estimate parameters with only aggregate data on market shares and some second-choice data? and some price-cost margins? (w/ Julie Mortimer and Paul Sarkis)

Not talking about first three today...

First: Some Review

Cournot Model (1838) / Nash in Quantities

Firms simultaneously choose q_f to maximize profits and calculate FOC's where $Q = \sum_{f=1}^{N} q_f$ the total output of the industry.

$$\pi_i(q_f) = P(Q) \cdot q_f - C_f(q_f)$$
$$\frac{\partial \pi_i(q_f)}{\partial q_f} = (P(Q) - C_f'(q_f)) + q_f \cdot P'(Q) \cdot \frac{\partial Q}{\partial q_f} = 0$$

Cournot competition implies that $\frac{\partial Q}{\partial q_f} = 1$ and $\frac{\partial q_g}{\partial q_f} = 0$ for $f \neq g$.

$$\underbrace{P(Q) + \overbrace{P'(Q) \cdot q_f}^{\text{Cournot Distortion}}}_{MR} = mc_f$$

We have that $\frac{q_f}{Q} \cdot \frac{\partial Q}{\partial q_f} = \frac{q_f}{\sum_{g=1}^{n} q_g} \equiv s_f$ or market share.

Asymmetric Cournot and HHI

- The Cournot markup / Lerner Index is just:

$$\frac{P - mc_f}{P} = \frac{s_f}{|\epsilon_D|}$$

$$\sum_{f=1}^{N} \frac{P - mc_f}{P} s_f = \sum_{f=1}^{n} \frac{s_f^2}{|\epsilon_D|}$$

- Cournot: markups are proportional to market-share.
- Obviously this nests symmetric case where $q_f = \frac{Q}{n}$ or $s_i = \frac{1}{n}$.
- $HHI = \sum_f s_f^2$ is proportional to the share-weighted average markup.
- Can also work backwards form HHI to get effective "number of firms".
- Here HHI is in units of $[0,1]$ instead of $[0,10000]$.

$$HHI = \sum_{i=1}^{N} s_i^2 = \frac{1}{n^*} \rightarrow n^* = \frac{1}{HHI}.$$

HHI and Welfare

Under Cournot (and only Cournot) with constant MC, we can relate HHI to particular measures of welfare:

- Cowling Waterson (1976) relate HHI to producer share of revenue:

$$HHI = \epsilon_d \cdot \frac{PS}{R}$$

- Spiegel (2020) relates HHI to producer share of surplus:

$$HHI = \frac{1}{\epsilon_d \left(Q*\right)} \cdot \frac{PS}{CS}$$
$$\frac{CS}{TS} = \frac{1}{1 + \epsilon_d \left(Q*\right) \cdot HHI}$$

Cournot: Discussion

- DOJ/FTC describe markets as:
  - Highly Concentrated: $HHI \geqslant 2500$ (4 firms)
  - Moderately Concentrated: $HHI \in [1500, 2500]$. $\Delta HHI \geqslant 250$ merits scrutiny.
  - Un-Concentrated: $HHI \leqslant 1500$. (> 6 firms)
- This has been important aspect of US HMG since 1980.
- But this makes market share or market definiton high stakes
- Whole Foods Case (2007)
  - FTC alleges 2:1 merger for "premium natural organic supermarket"
  - Merging parties argue they have combined market share less than 2% of supermarket market.
  - Do we want cases to hinge on merging parties or whether or not Wal-Mart and Amazon are "in market"?

Multiproduct (Asymmetric) Bertrand

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

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

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

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

$$\underbrace{p_j \cdot (1 + 1/\epsilon_{jj}(\mathbf{p}))}_{\text{MR}} = c_j + \sum_{k \in \mathcal{J}_f \setminus j} (p_k - c_k) \cdot D_{jk}(\mathbf{p})$$

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.

In Matrix/Vector Form

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

It is helpful to define the demand derivatives $\Delta_{(j,k)}(\mathbf{p}) = -\frac{\partial q_j}{\partial p_k}(\mathbf{p})$ and the ownership matrix:

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

We can re-write the FOC in matrix form where $\odot$ denotes Hadamard product (element-wise):

$$\begin{aligned} \mathbf{q}(\mathbf{p}) &= (\mathcal{H} \odot \Delta(\mathbf{p})) \cdot (\mathbf{p} - \mathbf{mc}), \\ \mathbf{mc} &= \mathbf{p} - \underbrace{(\mathcal{H} \odot \Delta(\mathbf{p}))^{-1} \mathbf{q}(\mathbf{p})}_{\eta(\mathbf{p},\mathbf{q},\theta_2)}. \end{aligned}$$

Solving this is easy (just inverting a matrix).

Unilateral Effects: What do mergers do?

Change the owenrship matrix !

$$\underbrace{p_j \cdot (1 + 1/\epsilon_{jj}(\mathbf{p}))}_{\text{MR}} = c_j + \sum_{k \in \mathcal{J}_f \setminus j} (p_k - c_k) \cdot D_{jk}(\mathbf{p})$$

- Upward Pricing Pressure: how does RHS change?

$$UPP_j = \Delta c_j + \sum_{k \in \mathcal{J}_g} (p_k - c_k) \cdot D_{jk}(\mathbf{p})$$

- This tells us whether or not opportunity costs increase/decrease?
- To get price predictions we have to solve J equations and J unknowns $\mathbf{p}$.
  - Need to know $\frac{\partial \mathbf{p}}{\partial \mathbf{c}}$ (pass-through matrix).
- What do we miss? Coordinated Effects, Dynamics, Innovation, etc.

Merger Simulation

What does a merger do? change the ownership matrix.

- Step 1: Recover marginal costs $\widehat{\mathbf{mc}} = \mathbf{p} + (\mathcal{H} \odot \Delta(\mathbf{p}))^{-1}\mathbf{q}(\mathbf{p})$.
- Step 1a: (Possibly) adjust marginal cost $\widehat{\mathbf{mc}} \cdot (1 - e)$ with some cost efficiency e.
- Step 2: Change the ownership matrix $\mathcal{H}^{pre} \to \mathcal{H}^{post}$.
- Step 3: Solve for $\mathbf{p}^{post}$ via: $\mathbf{p} = \widehat{\mathbf{mc}} - \Delta(\mathbf{p})^{-1}q(\mathbf{p})$.

- The first step is easy (just a matrix inverse).
- The second step is trivial.
- The third step is tricky because we have to solve an implicit system of equations. $\mathbf{p}$ is on both sides. (It turns out it isn't a contraction).

Merger Simulation

What does a merger do? change the ownership matrix.

- Step 1: Recover marginal costs $\widehat{\mathbf{mc}} = \mathbf{p} + (\mathscr{H} \odot \Delta(\mathbf{p}))^{-1}\mathbf{q}(\mathbf{p})$.
- Step 1a: (Possibly) adjust marginal cost $\widehat{\mathbf{mc}} \cdot (1-e)$ with some cost efficiency e.
- Step 2: Change the ownership matrix $\mathscr{H}^{pre} \rightarrow \mathscr{H}^{post}$.
- Step 3: Solve for $\mathbf{p}^{post}$ via: $\mathbf{p} = \widehat{\mathbf{mc}} - \Delta(\mathbf{p})^{-1}q(\mathbf{p})$.

- The first step is easy (just a matrix inverse).
- The second step is trivial.
- The third step is tricky because we have to solve an implicit system of equations. $\mathbf{p}$ is on both sides. (It turns out it isn't a contraction).

Solution Methods

How do we solve: $\mathbf{p} = \widehat{\mathbf{mc}} - \Delta(\mathbf{p})^{-1} q(\mathbf{p})$?

1. Gauss Jacobi: Simultaneous Best Reply $p_j^{k+1}(\mathbf{p^k_{-j}})$.
2. Gauss Seidel: Iterated Best Response $p_j^{k+1}(\mathbf{p^{k+1}_{<j}}, \mathbf{p^k_{>j}})$.
3. Newton's Method: Set $\mathbf{p} - \widehat{\mathbf{mc}} + \Delta(\mathbf{p})^{-1} q(\mathbf{p}) = 0$
   but requires derivatives of $\Delta(\mathbf{p})^{-1} q(\mathbf{p})$
4. Fixed point iteration: $\mathbf{p} \hookleftarrow \widehat{\mathbf{mc}} - \Delta(\mathbf{p})^{-1} q(\mathbf{p})$
   - Turns out this is not a contraction.
   - But you can get lucky... $\mathbf{p} - \widehat{\mathbf{mc}} + \Delta(\mathbf{p})^{-1} q(\mathbf{p}) = 0$ means you have satisfied FOC's
5. Alternative fixed point.

Exploit the logit formula

For the logit the Δ matrix (for a single market) looks like:

$$\Delta_{(j,k)}(\mathbf{p}) = \left\{ \begin{array}{cc} \int \alpha_i \cdot s_{ij} \cdot (1 - s_{ij})\, \partial F_i & \text{if } j = k \\ -\int \alpha_i \cdot s_{ij} \cdot s_{ik}\, \partial F_i & \text{if } j \neq k \end{array} \right\}$$

Which we can factor into two parts (for plain logit):

$$\Delta(\mathbf{p}) = \underbrace{\text{Diag}\left[\alpha\, \mathbf{s}(\mathbf{p})\right]}_{\Lambda(\mathbf{p})} - \underbrace{\alpha \cdot \mathbf{s}(\mathbf{p})\mathbf{s}(\mathbf{p})'}_{\Gamma(\mathbf{p})}$$

$\Gamma(\mathbf{p})$ 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.

Morrow Skerlos (2010) Fixed Point

- 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).
- Honestly, this is the best way to solve large pricing games. It nearly always wins and doesn't require derivatives.
- Coincidentally, this is what `PyBLP` defaults to.

Review: “Classic” BLP (1995) Models

Demand: Product Space

- Remember we need $\epsilon_{jj}(\mathbf{p})$ and $D_{jk}(\mathbf{p}) = \frac{\epsilon_{jk}(\mathbf{p})}{\epsilon_{jj}(\mathbf{p})} \cdot \frac{q_j}{q_k}$ so $\epsilon_{jk}(\mathbf{p})$.
- Multiple "markets" t with different prices/assortment.
- One approach might be constant elasticity or log-log

$$\log Q_{jt} = \beta_0 + \sum_k \beta_{j,k} \log P_{kt} + e_{jt}$$

- Even better might be Almost Ideal Demand (Deaton Muellbauer 1980) or trans-log

$$\log Q_{jt} = \beta_0 + \sum_k \beta_{j,k} \log P_{kt} + \sum_k \gamma_{j,k} \log P_{kt} \log P_{jt} + e_{jt}$$

- Problem: If we have J products, we need to estimate J^2 elasticities.
- Bigger Problem: every time P_k shows up on RHS, we need an instrument (!)

Demand: Product Space (Multinomial Probit)

- Assume consumers have unit demand and make discrete choices
- Label $j=0$ the no-purchase or outside option has $V_{i0t}=0$

$$u_{ijt} = V_{ijt} - \alpha p_{jt} + \varepsilon_{ijt} \text{ with } \boldsymbol{\varepsilon_{it}} \sim \mathcal{N}(0,\Omega)$$

- Idea: calculate $\sigma_{ijt} = \mathbb{P}(u_{ijt} > u_{ikt} \text{ for all } k \neq j)$
- But some problems...
  - Have to simulate draws from normals to calculate s_{ijt} (no closed form)
  - Still have J^2 elements in Ω (Maybe more like $(J+1) \cdot J/2$ with symmetry)
  - Still need J instruments.

Demand: Characteristic Space (Multinomial Logit)

- What if we assumed that products were just bundles of characteristics
- Assume that $V_{ijt} = x_{jt}\beta - \alpha p_{jt}$
- Also change the distribution of ε_{ij} to be Type I EV (Gumbel/Logit)

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \varepsilon_{ijt} \text{ with } \varepsilon_{ijt} \sim \text{IID Type I EV.}$$

- Now σ_{jt} is easy to calculate (and doesn't depend on i)

$$\sigma_{jt} \equiv \mathbb{P}(u_{ijt} > u_{ikt} \text{ for all } k \neq j) = \frac{e^{x_{jt}\beta - \alpha p_{jt}}}{1 + \sum_k e^{x_{kt}\beta - \alpha p_{kt}}}$$

- But now IV is hard: because $\mathbf{p_t}$ enters the model nonlinearly

Inversion: IIA Logit (Berry 1994)

Add unobservable error for each s_{jt} labeled ξ_{jt} and set
observed shares $\mathcal{S}_{jt} = \sigma_j(\boldsymbol{\delta_t}, \alpha, \beta)$ predicted shares.

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

$$\log \mathcal{S}_{jt} - \log \mathcal{S}_{0t} = \delta_{jt} - \underbrace{\delta_{0t}}_{=0} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

- 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 $\text{Corr}(\xi_{jt}, p_{jt}) \neq 0$
- 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).
- This is just Linear IV!

Multinomial Logit: Disadvantages

- To set observed shares = predicted shares we need large markets (lots of people making decisions) otherwise we have measurement error in shares.
- The logit has some unfortunate properties:
  - $\epsilon_{jj} = \alpha \cdot p_j \cdot (1 - s_j)$ (increasing in own price!)
  - $\epsilon_{jk} = \alpha \cdot p_j \cdot s_k$ (doesn't depend on s_j!)
  - IIA/Proportional substitution $\frac{s_j}{s_k}$ is unaffected by x_l
  - IIA/Proportional substitution $D_{j \to k} = \frac{s_k}{1-s_j}$

But we can solve these with unobserved heterogeneity or random effects.

$$u_{ijt} = \delta_{jt} + \mu_{ijt}(\theta_2) + \varepsilon_{ijt}.$$

where μ_{ijt} is mean zero.

Multinomial Logit: Disadvantages

- To set observed shares = predicted shares we need large markets (lots of people making decisions) otherwise we have measurement error in shares.
- The logit has some unfortunate properties:
  - $\epsilon_{jj} = \alpha \cdot p_j \cdot (1 - s_j)$ (increasing in own price!)
  - $\epsilon_{jk} = \alpha \cdot p_j \cdot s_k$ (doesn't depend on s_j!)
  - IIA/Proportional substitution $\frac{s_j}{s_k}$ is unaffected by x_l
  - IIA/Proportional substitution $D_{j \to k} = \frac{s_k}{1-s_j}$

But we can solve these with unobserved heterogeneity or random effects.

$$u_{ijt} = \delta_{jt} + \mu_{ijt}(\theta_2) + \varepsilon_{ijt}.$$

where μ_{ijt} is mean zero.

What about Nested Logit?

The most common solution allow for correlated preferences among products in the same nest

$$u_{ijt} = \delta_{jt} + \eta_{i,g(j),t}(\rho) + \varepsilon_{ijt}(\rho).$$

This model is only as good as our groupings $g(j)$.

- We still get IIA/proportional substitution within nest.
- Define: $Z\left(\rho, s_g\right) = [\rho + (1-\rho)s_g] \in (0,1]$
- For same group :

$$D^{*}_{j\to k} = \frac{s_{k|g}}{Z^{-1}\left(\rho, s_g\right) - s_{j|g}}$$

- For different groups :

$$D^{**}_{j\to k} = \frac{s_k(1-\rho)}{1 - s_{j|g} \cdot Z\left(\rho, s_{g(j)}\right)}$$

This gets used a ton at DOJ/FTC (with all goods in one nest). Why?

Inversion: Berry, Levinsohn, Pakes (1995)

We can't solve for δ_{jt} directly this time

$$\sigma_j(\boldsymbol{\delta_t}; \mathbf{x_t}; \theta_2) \quad = \quad \int \frac{\exp[\delta_{jt} + \mu_{ij}]}{1 + \sum_k \exp[\delta_{kt} + \mu_{ik}]} f(\boldsymbol{\mu_i}|\widetilde{\theta}_2)$$

- 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 $\widetilde{\theta}_2 = [\Pi, \Sigma]$ and $\theta_2 = [\alpha, \widetilde{\theta}_2]$
- 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.

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

  - 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].

BLP Pseudocode

From the outside, in:

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

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

- Inner Loop:
    - Solve for δ so that $\sigma_{jt}(\delta, \theta_2) = \tilde{S}_{jt}$.
        - Computing $s_{jt}(\delta, \theta)$ requires numerical integration (quadrature or monte carlo).
    - We can do IV-GMM to recover $\hat{\alpha}(\theta_2), \hat{\beta}(\theta_2), \hat{\xi}(\theta_2)$.

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

    - Use $\hat{\xi}(\theta_2)$ 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

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

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

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

BLP Extensions: Panel Data

- 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.

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

Recovering Marginal Costs

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

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

Which we can solve for ω_{jt}:

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

- $f(\cdot)$ 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).

Simultaneous Supply and Demand: in details

(a) For each market t: solve $\mathcal{S}_{jt} = \sigma_{jt}(\delta_{\cdot t}, \theta_2)$ 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 of Conlon Gortmaker (2020)

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

Micro Moments

Micro BLP is used a lot

Paper	Industry	Paper	Industry
Petrin (2002)	Automobiles	Barwick, Cao, & Li (2017)	Automobiles
Berry, Levinsohn, & Pakes (2004)	Automobiles	Murry (2017)	Automobiles
Thomadsen (2005)	Fast Food	Wollmann (2018)	Commercial Vehicles
Goeree (2008)	Personal Computers	S. Li (2018)	Automobiles
Ciliberto & Kuminoff (2010)	Cigarettes	Y. Li, Gordon, & Netzer (2018)	Digital Cameras
Nakamura & Zerom (2010)	Coffee	Backus, Conlon, & Sinkinson (2021)	Cereal
Beresteanu & Li (2011)	Automobiles	Grieco, Murry, & Yurukoglu (2021)	Automobiles
S. Li (2012)	Automobiles	Neilson (2021)	Primary Schools
Copeland (2014)	Automobiles	Armitage & Pinter (2022)	Automobiles
Starc (2014)	Health Insurance	Döpper, MacKay, Miller, & Stiebale (2022)	Retail
Ching, Hayashi, & Wang (2015)	Nursing Homes	Bodéré (2023)	Preschools
S. Li, Xiao, & Liu (2015)	Automobiles	Montag (2023)	Laundry Machines
Nurski & Verboven (2016)	Automobiles	Conlon & Rao (2023)	Distilled Spirits
⋮	⋮	⋮	⋮

- Many empirical IO papers use the "micro BLP" approach
  1. Impose the Berry, Levinsohn, & Pakes (1995) share constraint (unlike Grieco et al. 2023)
  2. Stack product-level or "aggregated" moments with "micro" moments from consumer surveys
  3. Make all this easy to do with PyBLP

A standardized framework (Conlon Gortmaker 2023)

- Aggregate data generated market-by-market t
  - Products $j \in \mathcal{J}_t$ have observed characteristics x_{jt}, unobserved quality ξ_{jt}
  - Consumer types $i \in \mathcal{I}_t$ have observed demographics y_{it}, unobserved preferences ν_{it}
  - Market shares $s_{jt} = \sum_i w_{it} s_{ijt}$ integrate over consumer mass, each type has known weight w_{it}

- Micro data generated dataset-by-dataset d, conditional on aggregate data
  - Results $\{(t_n, j_n, y_{i_n t_n})\}_{n \in \mathcal{N}_d}$ from independent surveys of selected consumers
  - Each consumer n was surveyed with known probability $w_{di_n j_n t_n}$

- Often only have or willing to use summary stats (cost, compatibility, interpretability, etc.)
  - Smooth functions $f(\overline{v}_d)$ of averages $\overline{v}_d = \frac{1}{N_d} \sum_n v_{di_n j_n t_n}$

- *"I want to match the mean demographic of consumers who purchased a product"*
  - "$\mathbb{E}[y_{it} \mid j \neq 0]$" $\leftarrow$ Let $w_{dijt} = 1\{j \neq 0\}$ and $v_{dijt} = y_{it}$

Standard micro moments

$$u_{ijt} = x'_{jt}(\beta_0 + \Pi_0 y_{it} + \Sigma_0 \nu_{it}) + \xi_{jt} + \varepsilon_{ijt}$$

- With only product-level aggregate data, often difficult to accurately estimate Π_0 and Σ_0
  - Often limited cross-market variation in demographic distributions and choice sets
- What within-market micro variation is informative about Π_0?
  - Literature tends to match stats that look like "$\mathbb{C}(x_{jt}, y_{it} \mid j \neq 0)$"
- What about Σ_0?
  - Literature emphasizes second choices, e.g. "$\mathbb{C}(x_{jt}, x_{k(\text{-}j)t} \mid j, k \neq 0)$"
- What about β_0?
  - Only indirectly: β_0 enters s_{ijt} only through $\delta_{jt} = x'_{jt}\beta_0 + \xi_{jt}$, pinned down by share constraint

Support for most cases

Paper	Micro moments shorthand	Paper	Micro moments shorthand
Petrin (2002)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[y_i \mid j \in \mathcal{J}]$	Barwick et al. (2017)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$
Berry et al. (2004)	$\mathbb{C}(x_j, y_i \mid j \neq 0)$, $\mathbb{C}(x_j, x_{k(-j)} \mid j, k \neq 0)$	Murry (2017)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Thomadsen (2005)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	Wollmann (2018)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Goeree (2008)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	S. Li (2018)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$
Ciliberto & Kuminoff (2010)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$	Y. Li et al. (2018)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$
Nakamura & Zerom (2010)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	Backus et al. (2021)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$, $\mathbb{C}(x_j, y_i \mid j \neq 0)$
Beresteanu & Li (2011)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	Grieco et al. (2021)	$\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$, $\mathbb{C}(x_j, x_{k(-j)} \mid j, k \neq 0)$
S. Li (2012)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[y_i \mid j \in \mathcal{J}]$	Neilson (2021)	$\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$
Copeland (2014)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$	Armitage & Pinter (2022)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Starc (2014)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$	Döpper et al. (2022)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Ching et al. (2015)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	Bodéré (2023)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$
S. Li et al. (2015)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	Montag (2023)	$\mathbb{C}(x_j, y_i \mid j \neq 0)$, $\mathbb{C}(x_j, x_{k(-j)} \mid j, k \neq 0)$
Nurski & Verboven (2016)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$, $\mathbb{C}(x_j, y_i \mid j \neq 0)$	Conlon & Rao (2023)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$, $\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Framework supports most cases we've seen
  - Demographic/choice-based sampling, conditioning, covariances, second choices $k \neq j$ too!

Micro BLP estimator: Tips

- Micro moments m match smooth functions $f_m(\cdot)$ of simple averages, called micro parts p

$$\bar{v}_p = \frac{1}{N_{d_p}} \sum_n v_{p i_n j_n t_n} \xrightarrow{P_{\mathrm{A}}} v_p(\theta_0) = \mathbb{E}_{\mathrm{A}}[v_{p i_n j_n t_n}] = \frac{\sum_t \sum_i \sum_j w_{it} s_{ijt}(\theta_0) w_{d_p ijt} v_{pijt}}{\sum_t \sum_i \sum_j w_{it} s_{ijt}(\theta_0) w_{d_p ijt}}$$

- Users define the weights w_{ijt}
  - Match only relevant markets t (ie: if survey is from 2020, don't match 2015 data!)
  - Subsets of products (only BMW's) and subsets of individuals (high income households with children).
- Most pairs of datasets have at least some incompatibilities in timing, variables, etc.
  - Optimal micro moments will only work well if incompatibilities are small
  - If large, match moments you expect to be compatible, e.g. correlations if scales are different
- Quadrature behaves poorly with discontinuities in moments like "$\mathbb{E}[x_{jt} \mid y_{it} < \bar{y},\ j \neq 0]$"
  - Instead, use Monte Carlo methods or moments continuous in y_{it} like "$\mathbb{C}(x_{jt}, y_{it} \mid j \neq 0)$"

Instruments and Identification

Exclusion Restrictions (see Berry Haile 2014)

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

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?

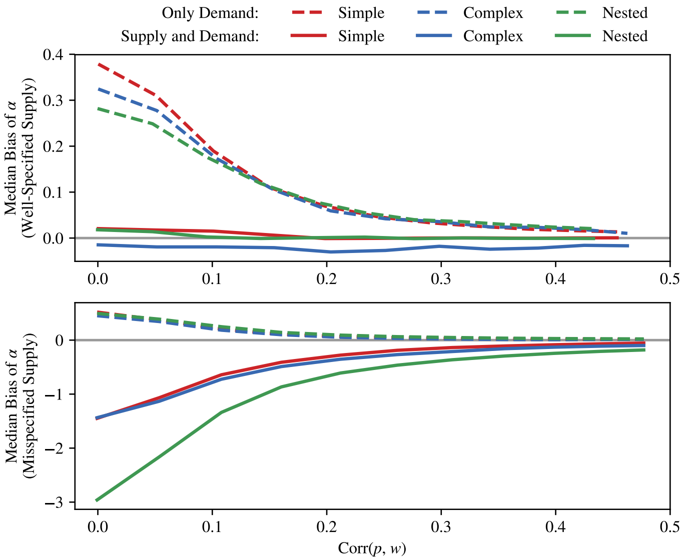
Parametric Identifcation

- 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.

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! $\eta_{jt}(\mathbf{p}, \mathbf{s}, \xi_t, \omega_t, \mathrm{x}_t, \mathrm{w}_t, \mathrm{v}_t, \mathrm{y}_t, \theta_2)$:

- 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.

Differentiation Instruments: Gandhi Houde (2019)

- 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?

Takeaway

What does this mean:

- We should always check first stage $\mathbb{E}[p \mid x, z]$ before we do anything else.
- 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. (but you need a good first stage)

Some New Things

Back to Diversion

Raise price of good j. People leave. What fraction of leavers switch to k?

$$D_{jk}(p_j, p_{-j}) = \frac{\frac{\partial q_k}{\partial p_j}}{\left|\frac{\partial q_j}{\partial p_j}\right|}$$

It's one of the best ways economists have to characterize competition among sellers.

- High Diversion: Close Substitutes $\rightarrow$ Mergers more likely to increase prices.
- Very low diversion $\rightarrow$ products may not be in the same market.
  (ie: Katz & Shapiro). This is just hypothetical monopolist or SSNIP test.
- Demand Derivatives NOT elasticities.
- No equilibrium responses.

Diversion: In Practice

1. Calculated from an estimated demand system (ratio of estimated cross-price to own-price demand derivatives)
2. Consumer surveys (what would you buy if not this?)
3. Obtained in 'course of business' (sales reps, internal reviews)

Antitrust authorities may prefer different measures in different settings. Are they concerned about:

- Small but widespread price hikes?
- Product discontinuations or changes to availability?

Is it sufficient to rely on data from merging firms only?

- Do we need diversion to other products in the 'market' or other functions of market-level data?
- Discrete-choice demand models imply that 'aggregate diversion' (including to an outside good) sums to one.

Diversion as Treatment Effects

Outcome $Y_i \in \{0,1\}$ denotes the event that consumer i purchases product k: $d_{ik}(P_j) = 1$.

Treatment $T_i \in \{0,1\}$ denotes the event that consumer i does **not** purchase product j. In other words $T_i = 0$ implies $d_{ij}(P_j) = 1$ and $T_i = 1$ implies $d_{ij}(P_j) = 0$.

Instrument $Z_i = P_j$ the price of j induces consumers into not purchasing j.

Theorem (Conlon Mortimer (RJE 2021))

Under the following conditions:

(a) Mutually Exclusive and Exhaustive Discrete Choice: $d_{ij} \in \{0,1\}$ *and* $\sum_{j \in \mathcal{J}} d_{ij} = 1$.

(b) Exclusion: $u_{ik}(p_j, x) = u_{ik}(p'_j, x)$ *for all* $k \neq j$ *and any* (p_j, p'_j)*;*

(c) Monotonicity: $u_{ij}(p'_j, x) \leqslant u_{ij}(p_j, x)$ *for all* i *and any* $(p'_j > p_j)$*; and*

(d) Existence of a first-stage: $Pr(d_{ij}(p_j, x) = 0) \neq Pr(d_{ij}(p'_j, x) = 0)$ *for* $(p'_j > p_j)$*;*

(e) Random Assignment: $(u_{ij}(P_j, x), u_{ik}(P_j, x)) \perp P_j$.

then the Wald estimator:

$$\frac{q_k(p'_j, x) - q_k(p_j, x)}{-\left(q_j(p'_j, x) - q_j(p_j, x)\right)} = \mathbb{E}[D_{jk,i}(x) | d_{ij}(p_j, x) > d_{ij}(p'_j, x)]$$

Diversion as MTE

What can we learn from a change in p_j?

$$\frac{q_k(p_j',x)-q_k(p_j,x)}{-\Big(q_j(p_j',x)-q_j(p_j,x)\Big)} = \mathbb{E}[D_{jk,i}(x)|d_{ij}(p_j,x)>d_{ij}(p_j',x)]$$

$$= \int D_{jk,i}(x)\, w_i(z_j,z_j',x)\,\partial F_i \quad \text{with } w_i(z_j,z_j',x) = \frac{q_{ij}(z_j,x)-q_{ij}(z_j',x)}{q_j(z_j,x)-q_j(z_j',x)}$$

- Individual diversion ratios in logit family are $D_{jk,i} = \frac{s_{ik}}{1-s_{ij}}$ and don't depend on p_j, z_j
- Which individuals respond to z_j or p_j determines the weighting scheme only.

Weights

Again $D_{jk,i} = \frac{s_{ik}}{1-s_{ij}}$ for logit family and plain logit $D_{jk,i} = \frac{s_k}{1-s_j}$ for all i and weights don't matter!

$$\int D_{jk,i}(x)\, w_i(z_j, z_j', x)\, \partial F_i$$

	$w_i(z_j, z_j', x) \propto$
second choice data	$s_{ij}(x)$
price change $\frac{\partial}{\partial p_j}$	$s_{ij}(x) \cdot (1 - s_{ij}(x)) \cdot \lvert\alpha_i\rvert$
characteristic change $\frac{\partial}{\partial x_j}$	$s_{ij}(x) \cdot (1 - s_{ij}(x)) \cdot \lvert\beta_i\rvert$
small quality change $\frac{\partial}{\partial \xi_j}$	$s_{ij}(x) \cdot (1 - s_{ij}(x))$

A Useful (?) Extension (Conlon, Mortimer, Sarkis 2023)

Definitions: Second Choices

For any (semiparametric) mixture of logits we can write the probability that individual i chooses k as their second choice given that j is their first choice as:

$$
\begin{aligned}
D_{j \to k} &\equiv \mathbb{P}(\text{ chooses } k \in \mathcal{J} \backslash \{j\} \mid \text{ chooses } j \in \mathcal{J}) \\
&= \sum_{i=1}^{I} \pi_i \cdot \frac{s_{ik}}{1 - s_{ik}} \cdot \frac{s_{ij}}{s_j}
\end{aligned}
$$

It is convenient to interpret $D_{j \to k}$ as the (j, k)th entry in the second-choice matrix $\mathbf{D}(\mathbf{S}, \pi)$.

Our Semiparametric Problem

$$\min_{(\mathbf{S},\pi)\geqslant 0} \|\mathscr{P}_\Omega(\mathscr{D} - \mathbf{D}(\mathbf{S},\pi))\|_{\ell_2} + \lambda \|\mathcal{S} - \mathbf{S}\,\pi\|_{\ell_2} \text{ with } \|\pi\|_{\ell_1} \leqslant 1, \quad \|\mathbf{s_i}\|_{\ell_1} \leqslant 1.$$

- Constraints: Choice probabilities s_{ij} sum to one, type weights π_i sum to one.
- Use cross validation to select $\#$ of types I and Lagrange multiplier λ.
- Not-convex but not very difficult either.
- ℓ_1 constraints lead to sparsity.
- Goal: estimate $\mathbf{s_i}$ (choice probabilities) and corresponding weights π_i (Finite Mixture)
- Can also match price cost margins

Would this be useful to folks at CMA?

Second Choice Matrix

- Individual i's share for each choice given by $\mathbf{s_i} = [s_{i0}, s_{i1}, \ldots, s_{iJ}]$.
- Aggregate shares by $\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} = \mathbf{s}$.
- The matrix of individual diversion ratios is given by $\mathbf{D}_i = \mathbf{s_i} \cdot \left[\frac{1}{(1-\mathbf{s_i})}\right]^T$.

We write the $(J+1) \times (J+1)$ matrix of second-choice probabilities as:

$$\mathbf{D} = \left(\sum_{i=1}^{I} \pi_i \cdot \mathbf{s_i} \cdot \left[\frac{1}{(1-\mathbf{s_i})}\right]^T \cdot \text{diag}(\mathbf{s_i}/\mathbf{s})^{-1}\right)^T$$

$$= \text{diag}(\mathbf{s})^{-1} \cdot \left(\sum_{i=1}^{I} \pi_i \cdot \left[\frac{\mathbf{s_i}}{(1-\mathbf{s_i})}\right] \cdot \mathbf{s_i}^T\right)$$

Second Choice Matrix: Continued

Under relatively general conditions, second-choice probabilities can be written as:

$$\mathbf{D} = \text{diag}(\mathbf{s})^{-1} \cdot \left(\sum_{i=1}^{I} \pi_i \cdot \begin{bmatrix} | \\ \mathbf{s_i} \\ | \end{bmatrix} \cdot \begin{bmatrix} - & \frac{\mathbf{s_i}}{1-\mathbf{s_i}} & - \end{bmatrix} \right)$$

- Each individual diversion ratio is of rank one since it is the outer product of $\mathbf{s_i}$ with itself (and some diagonal "weights").
- The (unrestricted) matrix of diversion ratios $\mathbf{D}$ is $(J+1) \times (J+1)$.
- Logit restricts $\mathbf{D}$ to be of rank one. Nested logit of rank $\leqslant G$ (the number of non-singleton nests). Mixed logit to $\text{rank}(\mathbf{D}) \leqslant I$ (but bound is likely uninformative).

In practice this works shockingly well (with some caveats).

References i

Armitage, S., & Pinter, F. (2022). Regulatory mandates and electric vehicle product variety.

Backus, M., Conlon, C., & Sinkinson, M. (2021). Common ownership and competition in the ready-to-eat cereal industry.

Barwick, P. J., Cao, S., & Li, S. (2017). Local protectionism, market structure, and social welfare: China's automobile market.

Beresteanu, A., & Li, S. (2011). Gasoline prices, government support, and the demand for hybrid vehicles in the United States. *International Economic Review*, *52*(1), 161–182.

Berry, S. T., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Berry, S. T., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, *112*(1), 68–105.

Bodéré, P. (2023). Dynamic spatial competition in early education: An equilibrium analysis of the preschool market in Pennsylvania.

Ching, A. T., Hayashi, F., & Wang, H. (2015). Quantifying the impacts of limited supply: The case of nursing homes. *International Economic Review*, *56*(4), 1291–1322.

Ciliberto, F., & Kuminoff, N. V. (2010). Public policy and market competition: how the master settlement agreement changed the cigarette industry. *The B.E. Journal of Economic Analysis & Policy*, *10*(1), Article 63.

Conlon, C., & Rao, N. (2023). Market power as second-worst regulation: Welfare consequences of post and hold pricing in distilled spirits.

Copeland, A. (2014). Intertemporal substitution and new car purchases. *RAND Journal of Economics*, *45*(3), 624–644.

Döpper, H., MacKay, A., Miller, N., & Stiebale, J. (2022). Rising markups and the role of consumer preferences.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, *76*(5), 1017–1074.

Grieco, P. L. E., Murry, C., Pinkse, J., & Sagl, S. (2023). Conformant and efficient estimation of discrete choice demand models.

Grieco, P. L. E., Murry, C., & Yurukoglu, A. (2021). The evolution of market power in the US auto industry.

Li, S. (2012). Traffic safety and vehicle choice: quantifying the effects of the 'arms race' on American roads. *Journal of Applied Econometrics*, *27*(1), 34–62.

Li, S. (2018). Better lucky than rich? Welfare analysis of automobile licence allocations in Beijing and Shanghai. *The Review of Economic Studies*, *85*(4), 2389–2428.

Li, S., Xiao, J., & Liu, Y. (2015). The price evolution in China's automobile market. *Journal of Economics & Management Strategy*, *24*(4), 786–810.

Li, Y., Gordon, B. R., & Netzer, O. (2018). An empirical study of national vs. local pricing by chain stores under competition. *Marketing Science*, *37*(5), 812–837.

Montag, F. (2023). Mergers, foreign competition, and jobs: Evidence from the US appliance industry.

Murry, C. (2017). Advertising in vertical relationships: An equilibrium model of the automobile industry.

Nakamura, E., & Zerom, D. (2010). Accounting for incomplete pass-through. *The Review of Economic Studies*, *77*(3), 1192–1230.

Neilson, C. (2021). Targeted vouchers, competition among schools, and the academic achievement of poor students.

Nurski, L., & Verboven, F. (2016). Exclusive dealing as a barrier to entry? Evidence from automobiles. *The Review of Economic Studies*, *83*(3), 1156–1188.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, *110*(4), 705–729.

References v

Starc, A. (2014). Insurer pricing and consumer welfare: Evidence from Medigap. *RAND Journal of Economics*, *45*(1), 198–220.

Thomadsen, R. (2005). The effect of ownership structure on prices in geographically differentiated industries. *RAND Journal of Economics*, *36*(4), 908–929.

Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, *108*(6), 1364–1406.