30 Years of BLP...

Chris Conlon

IIOC 2024

NYU Stern School of Business and NBER

Thank Organizers and Steve!

Thank Coauthors

- w/ Julie Mortimer
  - *Demand Estimation Under Incomplete Product Availability*
  - *Empirical Properties of Diversion Ratios*
  - w/ Paul Sarkis *Estimating Preferences and Substitution Patterns from Second-Choice Data Alone*
- w/ Nirupama Rao
  - *The Cost of Curbing Externalities with Market Power: Alcohol Regulations and Tax Alternatives*
- w/ Matt Backus and Michael Sinkinson
  - *Common Ownership and Competition in the Ready-To-Eat Cereal Industry*
- w/ Jeff Gortmaker
  - *Best Practices for Demand Estimation with pyBLP*
  - *Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP*

Organization

Past Lots of good ideas in "original" BLP95/99; including some key ones that got ignored and hopefully rediscovered.

- BLP was at its core about simultaneous supply and demand.
- Much like Chamberlain (1987) the section on optimal IV was ahead of its time.

Present Data and Computers are way better than in 1995

- Especially Micro Data/ Mini case study
- Trying to cram everything in the BLP box is not always the best idea.
  - Some BLP alternatives: analytic inverses, approximations, etc.

Future Can we realize the dream of non-parametric identification in estimation?

- Doing ML not just measuring its impact!

The“Classics” #1: Supply Side

Supply Side

Consider the multi-product Bertrand FOCs where $\arg\max_{p \in \mathcal{J}_f} \pi_f(\mathbf{p})$:

$$\pi_f(\mathbf{p}) \equiv \sum_{j \in \mathcal{J}_f} (p_j - c_j) \cdot \sigma_j(\mathbf{p}) + \sum_{k \in \mathcal{J}_g} (p_k - c_k) \cdot \sigma_k(\mathbf{p})$$

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

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

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

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

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

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

What's the point?

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

Demand systems have two main deliverables:

- Own-price elasticities $\epsilon_{jj}(\mathbf{p})$
- Substitution patterns
  - Cross elasticities: $\epsilon_{jk}(\mathbf{p}) = \frac{p_j}{q_k} \cdot \frac{\partial q_k}{\partial p_j}$
  - 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.

We did Nash-in-Prices because it is popular but we could have done something else.

Constructing Supply Moments

If we are willing to impose $MR = MC$ (as in the original BLP papers) we can recover implied markups/ marginal costs:

$$\mathbf{mc}(\theta_2) \equiv \mathbf{p} - \boldsymbol{\eta}(\mathcal{S}_t, \mathbf{p}_t, \chi_t, y_t, ; \theta_2)$$

$$f\left(\mathbf{p} - \boldsymbol{\eta}(\mathcal{S}_t, \mathbf{p}_t, \chi_t, y_t, ; \theta_2)\right) = [\mathrm{x}_{jt}, \mathrm{w}_{jt}]\theta_3 + g(q_{jt}) + \omega_{jt}$$

- $f(\cdot)$ is usually $\log(\cdot)$ or identity; it is actually a production function
- $g(q_{jt})$ captures returns to scale and requires an additional instrument

Simultaneous Supply and Demand

$$\sigma_j^{-1}(\mathcal{S}_t, \mathbf{p}_t, \chi_t, y_t; \theta_2) = [\mathrm{x}_{jt}\,, \mathrm{v}_{jt}]\,\theta_1 - \alpha p_{jt} + \xi_{jt}$$
$$f(p_{jt} - \eta_{jt}(\mathcal{S}_t, \mathbf{p}_t, \chi_t, y_t; \theta_2)) = [\mathrm{x}_{jt}\,, \mathrm{w}_{jt}]\,\theta_3 + \omega_{jt}$$

We can now form two sets of moments: $\mathbb{E}[\omega_{jt} \mid z_{jt}^s] = 0$ and $\mathbb{E}[\xi_{jt} \mid z_{jt}^d] = 0$

- These provide overidentifying restrictions for (θ_2, α)
- Conditional on θ_2 (distribution of random coefficients) and α this is just linear IV-GMM again.
- The derivatives $\left(\frac{\partial \xi_{jt}}{\partial \theta_2}, \frac{\partial \omega_{jt}}{\partial \theta_2}\right)$ beacuse of $\frac{\partial \eta_{jt}}{\partial \theta_2}$ in particular, are complicated (But `PyBLP` knows how to do these).
- As Steve has made clear this is likely a many weak IV situation many potential IV's (others $x_{-j}, w_{-j}, v_{-j}, y_t$), but hard to know which are strong.

The“Classics”#2: Optimal IV

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:

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

Problems: we don't know θ_0 and endogeneity.

Chamberlain (1987)

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

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

$$\xi_{jt} = \sigma_j^{-1}(\cdot, \theta_2) - [\mathrm{x}_{jt}, \mathrm{v}_{jt}]\,\theta_1 + \alpha p_{jt}$$

$$\omega_{jt} = f(p_{jt} - \eta_{jt}(\cdot, \theta_2)) - [\mathrm{x}_{jt}, \mathrm{w}_{jt}]\,\theta_3$$

For the exogenous variables: $\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \theta_1} \mid z_{jt}^d\right] = [\mathrm{x}_{jt}, \mathrm{v}_{jt}]$ and $\mathbb{E}\left[\frac{\partial \omega_{jt}}{\partial \theta_3} \mid z_{jt}^s\right] = [\mathrm{x}_{jt}, \mathrm{w}_{jt}]$.

For the endogenous prices: $\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \alpha} \mid z_{jt}^d\right] = \mathbb{E}[p_{jt} \mid z_{jt}^d]$ and $\mathbb{E}\left[\frac{\partial \omega_{jt}}{\partial \alpha} \mid z_{jt}^s\right] = \mathbb{E}[f'(\cdot)(p_{jt} - \frac{\partial \eta_{jt}}{\partial \alpha}) \mid z_{jt}^s]$.

For the endogenous θ_2: $\mathbb{E}\left[\frac{\partial \boldsymbol{\xi}_t}{\partial \theta_2} \mid \mathbf{z_t^d}\right] = \mathbb{E}\left[\left[\frac{\partial \boldsymbol{\sigma}_t}{\partial \boldsymbol{\xi}_t}\right]^{-1}\left[\frac{\partial \boldsymbol{\sigma}_t}{\partial \theta_2}\right] \mid \mathbf{z_t^d}\right]$ and $\mathbb{E}\left[\frac{\partial \omega_{jt}}{\partial \theta_2} \mid \mathbf{z_t^s}\right]$

(but you can't condition on p_{jt})

Optimal Instruments

Even with an intitial guess of $\widehat{\theta}$, we still have that p_{jt} or η_{jt} depends on (ω_j, ξ_t) in a highly nonlinear way (no explicit solution!). But we have some options:

- Pray to the God of Sieves:
  - 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.
- Plug in a guess for first stage p_{jt}:
  - Reynaert Verboven (2014) suggest $\mathbb{E}[p_{jt} \mid \mathrm{x}_{jt}, \mathrm{w}_{jt}] = mc_{jt}$ (perfect competition), but might as well include other z_{jt}^d (like BLP instruments).
  - $\mathbb{E}[p_{jt} \mid z_{jt}^d]$ is easy, and non-parametric regression is pretty good.
- Use the nonlinearity in the model! (BLP 199)

Feasible Recipe (BLP 1999)

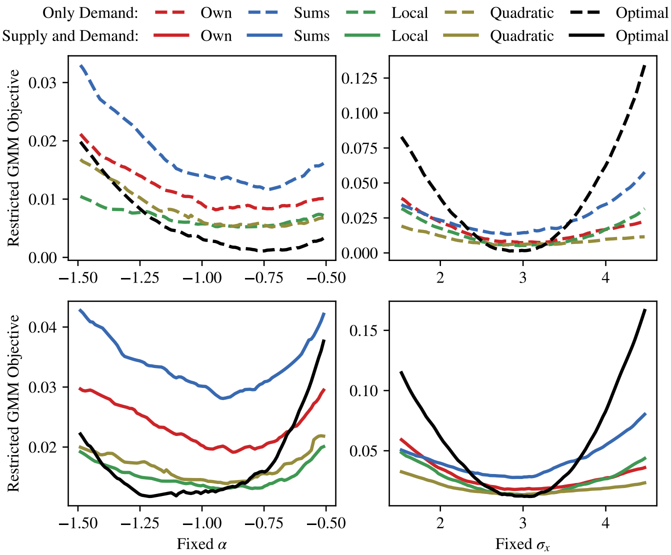
1. Fix $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ and draw $(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*)$ from empirical density
2. Solve firm FOC's for $\widehat{\mathbf{p}}_\mathbf{t}(\boldsymbol{\xi}^*, \boldsymbol{\omega}^*, \widehat{\theta})$ and shares $\mathbf{s}_\mathbf{t}(\widehat{\mathbf{p}}_\mathbf{t}, \widehat{\theta})$
3. Compute necessary Jacobian
4. 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.
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

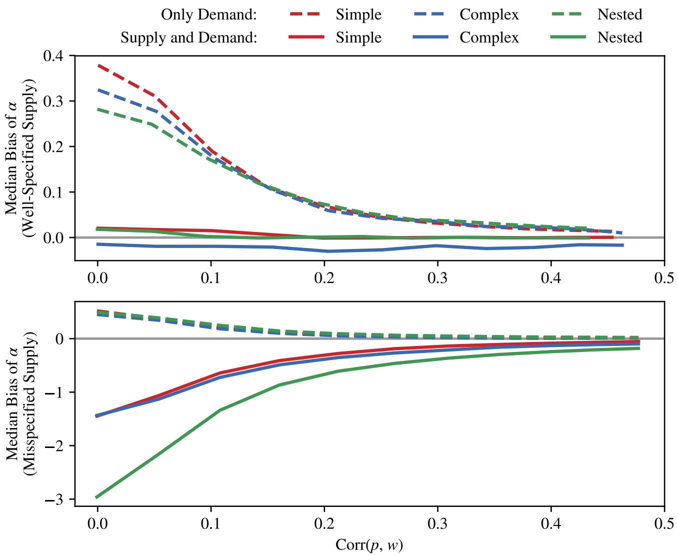
IV Comparison: Conlon and Gortmaker (2020)

				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)



Cost Shifters Really Matter Conlon Gortmaker (2020)



Aside: Optimal IV Everywhere! (Backus, Conlon, Sinkinson)

In our paper on testing conduct we are interested in testing $H_0 : \tau = 1$ and $H_a : \tau_= 0$

$$\omega_{jt} = p_{jt} - \tau \cdot \eta_{jt}^{(a)} - (1-\tau) \cdot \eta_{jt}^{(b)} - h(\mathrm{x}_{jt}, \mathrm{w}_{jt}; \theta_3)$$

The key to the test is to realize that optimal IV: $\mathbb{E}\left[\frac{\partial \omega_{jt}}{\partial \tau} \mid z_{jt}^s\right] = \mathbb{E}\left[\eta_{jt}^{(a)} - \eta_{jt}^{(b)} \mid z_{jt}^s\right]$.

- Instruments predict the difference in markups!
- Can run one non-parametric regression for $\mathbb{E}\left[\eta_{jt}^{(a)} - \eta_{jt}^{(b)} \mid z_{jt}^s\right]$ and another for the nuisance function (observed markup shifters) $h(\cdot)$.
- This is an easy way to do Berry Haile (2014). Duarte, Magnolfi, Sølvsten, Sullivan (2024) get a similar expression with a different approach.

Takeaway

What does this mean:

- Optimal IV aren't magic, you probably need good cost shifters.
- We should always check $\mathbb{E}[\mathbf{p} \mid \mathbf{z}]$ 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`.

BLP Today

The Good News

BLP today mostly works

- There was some concern (see Knittel Metaxoglou 2014) that BLP estimation was a bit "fragile"
- PyBLP knows how to take derivatives, and we spent a lot of time tuning default options: solving for shares, integration rules, optimization, etc.
- Most problems people have today are: too many parameters, not enough instruments (or really bad ones).
- Cross-market variation in number of products, or characteristics of products would also help.

The Bad News: Substitution

Your BLP model probably isn't as flexible as you think (Sorry!)

- The unrestricted matrix of of $D_{jk}(\mathbf{p})$'s (or elasticities) is $J \times J$ and probably not feasible to estimate directly. $\rightarrow$ need some dimension reduction.
- Logit simply assumes proportional substitution $D_{jk} = \frac{s_k}{1-s_j}$ so that $\mathbf{D}(\mathbf{p})$ is of rank one!
- The BLP solution is to project $\mathbf{D}(\mathbf{p})$ onto a lower-dimensional basis of x_j characteristics.
    - Ultimately the basis will be only as good as the characteristics x_{jt} with heterogeneous coefficients.
    - Distributional assumptions on $f(\beta_i)$ (ie: independent normal) further restrict the basis.
- The hardest thing to match is typically substitution to closest substitutes.
    - I worry that most BLP models (one RC, etc.) look too much like the plain logit.
- People often assume HUGE outside good shares and then have $D_{j0} > 0.9$.
    - We make fun of macro/trade for monopolistic competition but easy to estimate something close.

The Bad News: Elasticities

If you aren't careful, can still get some weird economics...

- Plain logit assumes in elasticities increasing in p_j and markups that are decreasing.
  - One RC or putting all products in one nest probably isn't going to fix that
- Simple versions tend to lead to high income / price insensitive consumers buying all the cars, yogurt, etc.
  - $\rightarrow$ Probably everyone should have a correlated RC on Price and Constant terms.
- How $f(y_i - \alpha_i p_j)$ looks as it may determine pass-through
  Griffith, Nesheim O'Connell (2018), Birchall Verboven (2022), Miravete, Seim, Thurk (2023)
  - $\rightarrow$ Lognormal α_i is probably a good idea
  - $\rightarrow$ bins for $\alpha(y_i)$ by income also good (break distributional relationship between p_{jt} and y_{it}).

Adding Micro Data

Micro BLP is used a lot: Probably you should use it to.

- Take advantage of much better data available in 2023!
- Stack product-level or "aggregated" moments with "micro" moments from surveys
  - Much easier to learn about interactions between demographics and characteristics.
  - Enables credible distributional analysis of policies.
- For demographic interactions Π, you want moments like $Cov(x_j, y_i) = \mathbb{E}[x_j \cdot y_i] - \mathbb{E}[x_j] \cdot \mathbb{E}[y_i]$
  - Often easier to construct $\mathbb{E}[x_j \mid y_i]$ or $\mathbb{E}[y_i \mid x_j]$ (often conditional on purchase $j \neq 0$)
- For normally distributed Σ, you want data on second choices.
  - Original MicroBLP used: $\mathbb{C}(x_j, x_{k(\text{-}j)} \mid j, k \neq 0)$
- The tricky bit is typically working out covariances and weighting matrix
  $\rightarrow$ this is what Conlon Gortmaker (2023) does! (with help from Myojo & Kanazawa (2012)).

Conlon Gormaker (2023): “Micro” PyBLP

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, Cao, & Li (2017)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$
Berry et al. (2004)	$\mathbb{C}(x_j, y_i \mid j \neq 0), \mathbb{C}(x_j, x_{k(\text{-}j)} \mid j, k \neq 0)$	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, Gordon, & Netzer (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, Conlon, & Sinkinson (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, Murry, & Yurukoglu (2021)	$\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0], \mathbb{C}(x_j, x_{k(\text{-}j)} \mid j,$
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, MacKay, Miller, & Stiebale (2022)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Ching, Hayashi, & Wang (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, Xiao, & Liu (2015)	$\mathbb{P}(j \in \mathcal{J} \mid i \in \mathcal{I})$	Montag (2023)	$\mathbb{C}(x_j, y_i \mid j \neq 0), \mathbb{C}(x_j, x_{k(\text{-}j)} \mid j, k$
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)	$\mathbb{E}[y_i \mid j \in \mathcal{J}], \mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$

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

Some words of caution

- There are optimal micro moments which approximate scores (see Conlon Gortmaker 2023).
  - There are no efficiency guarantees for inconsistent pilot estimates $\hat{\theta}$
  - For first step, can use standard moments or score at informed guess of θ_0
- 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
  - Problem exists for “typical moments” in Alcohol: $\mathbb{E}[Purch \mid y_{it}]$ vs. $\mathbb{E}[y_{it} \mid Purch]$ when $\mathbb{E}[Purch]$ is incompatible.
- Quadrature behaves poorly with discontinuities in moments like “$\mathbb{E}[x_{jt} \mid y_{it} < \overline{y}, j \neq 0]$”
  - Instead, use Monte Carlo methods or moments continuous in y_{it} like “$\mathbb{C}(x_{jt}, y_{it} \mid j \neq 0)$”

Complete Micro Data: Grieco Murry Pinkse Sagl (2023)

Like a one-step version of Goolsbee Petrin (2004)

$$(\hat{\beta}, \hat{\theta}_2, \hat{\delta}) = \arg \min_{\beta, \theta_2, \delta} - \log \hat{L}(\theta_2, \delta) + \hat{\Pi}(\beta, \delta)$$

1. the Mixed Data Likelihood Estimator:

$$\log \hat{L}(\theta, \delta) = \sum_{m=1}^{M} \sum_{j=0}^{J_m} \sum_{i=1}^{N_m} d_{ijm} (D_{im} \underbrace{\log \pi_{jm}^{y_{im}}(\theta_2, \delta)}_{\text{individual choices}} + (1 - D_{im}) \underbrace{\log \pi_{jm}(\theta_2, \delta)}_{\text{aggregate share}})$$

2. including the BLP moment conditions

$$\hat{\Pi}(\beta, \delta) = \frac{1}{2} \hat{g}^{\intercal}(\beta, \delta) \cdot \hat{\mathcal{W}} \cdot \hat{g}(\beta, \delta)$$

   where $\hat{g}^{\intercal}(\beta, \delta) = \sum_{m=1}^{M} \sum_{j=1}^{J_m} z_{jm} (\delta_{jm} - \beta^{\intercal} x_{jm})$.

Trick in the paper: getting $\hat{\mathcal{W}}$ correct.

Quick Case Studies: Micro Data, Second Choices, Supply

Grieco, Murry, Yurukoglu (QJE 2024)

			Demographic inte			
	β	σ	Income	Inc. sq.	Age	Rural
Price	−3.112	—	0.094	−0.462	2.065	—
	(1.124)		(0.010)	(0.133)	(0.122)	
Van	−7.614	5.538	—	—	—	—
	(0.598)	(0.133)				
SUV	−0.079	3.617	—	—	—	—
	(0.339)	(0.087)				
Truck	−7.463	6.309	—	—	—	3.007
	(0.898)	(0.310)				(0.340)
Footprint	0.534	1.873	—	—	—	—
	(0.261)	(0.118)				
Horsepower	1.018	1.246	—	—	—	—
	(0.954)	(0.361)				
Miles/Gal.	−0.965	1.645	—	—	—	—
	(0.211)	(0.151)				
Luxury	—	2.624	—	—	—	—
		(0.047)				
Sport	−3.046	2.617	—	—	—	—
	(0.549)	(0.075)				
EV	−5.549	3.798	—	—	—	—
	(1.406)	(0.511)				
Euro. brand	—	1.921	—	—	—	—
		(0.054)				
U.S. brand	—	2.141	—	—	—	—
		(0.048)				
Constant	—	—	0.362	—	—	—
			(0.034)			

Best MicroBLP application?

- Average characteristics by income, age, family size: $\mathbb{E}[x_j \mid i \in I, j \neq 0]$
- Covariance of characteristics for 1st and 2nd choices: $\mathbb{C}(x_j, x_{k(-j)} \mid j, k \neq 0)$
- I would make Price lognormal and put RC on constant...

Supply and Micro Data: (Conlon Rao 2023)

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

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

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

- 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: Lognormal
  - Constant β_i^0 (Overall demand for spirits)
  - Package Size: β_i^{1750} (Large vs. small bottles)

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

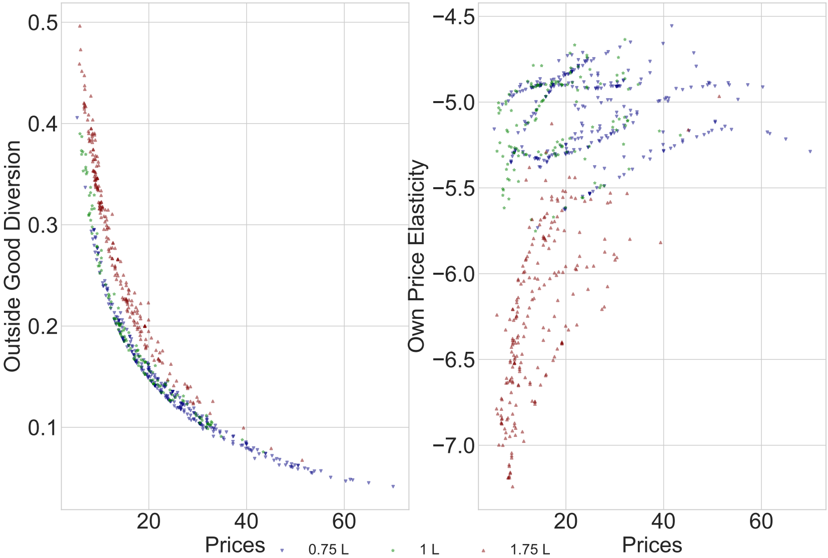
Π	Const	Price	1750mL
Below \$25k	2.928	-0.260	0.543
	(0.233)	(0.056)	(0.075)
\$25k-\$45k	0.184	-0.170	0.536
	(0.236)	(0.054)	(0.083)
\$45k-\$70k	0.000	-0.179	0.980
	(0.000)	(0.053)	(0.093)
\$70k-\$100k	-0.452	-0.496	0.608
	(0.227)	(0.051)	(0.079)
Above \$100k	-1.777	-1.543	0.145
	(0.234)	(0.047)	(0.055)
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).

$$\mathbb{E}[\omega_{jt}] = 0, \text{ with } \omega_{jt} = \left(p^w_{jt} - p^m_{jt} - \tau_{jt}\right) - \eta_{jt}\left(\theta_2\right).$$

- Match repeat purchase share within (Vodka, Gin, Rum, etc.) as in Atalay, Frost, Sorensen, Sullivan, and Zhu (2023)
- 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

Approximating the Problem

BLP 1995/1999 and Berry Haile (2014)

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

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

- 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: $\sigma_j^{-1}(\mathcal{S}_{\cdot t}, \rho) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} - \sum_{d=1}^{2} \rho_d \ln \left(\frac{\mathfrak{s}_{jt}}{\mathfrak{s}_{d(j),t}} \right)$ (Verboven 1996)
  - IPDL: $\sigma_j^{-1}(\mathcal{S}_{\cdot t}, \rho) = \ln \mathfrak{s}_{jt} - \ln \mathfrak{s}_{0t} - \mathbf{x}_{jt}\boldsymbol{\beta} - \alpha p_{jt} - \sum_{d=1}^{D} \rho_d \ln \left(\frac{\mathfrak{s}_{jt}}{\mathfrak{s}_{d(j),t}} \right) + \xi_{jt}$ (Fosgerau, Monardo, De Palma (2022))
- Anything with a share requires an IV (otherwise $\rho \to 1$).

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:

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

- $m_t^y = \sum_{i \in \mathcal{I}_t} w_{it} \cdot y_{it}$ 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} (endogenous shares!)

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:

$$x_{jt}^2 + \underbrace{\frac{1}{J} \cdot \sum_k x_{kt}^2}_{\text{constant for } t} - 2x_{jt} \cdot \sum_k \frac{1}{J} \cdot x_{kt}$$

and the artificial regressor is

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

- 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...

The nonparametric/machine learning future...?

What do you mean by non-parametric?

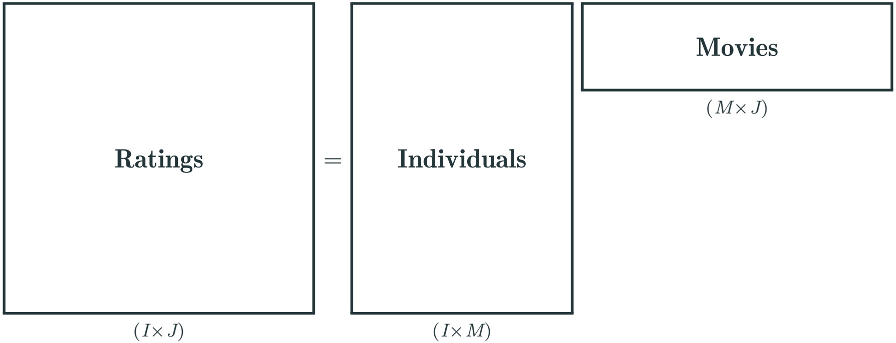
Mostly we mean putting a flexible distribution on $f(\beta_i, \alpha_i \mid \theta)$ (and keeping logit error on ε_{ij})

$$u_{ij} = \beta_i x_j - \alpha_i p_j + \xi_j + \varepsilon_{ij} \text{ with } f(\beta_i, \alpha_i \mid \theta)$$

- Compiani (QE 2022): approximate $\sigma_j^{-1}(\mathcal{S}_t, \mathbf{x}_t^{(2)})$ directly with Bernstein Polynomials (ditches $\varepsilon \rightarrow$ very hard)
- Ao Wang (JE 2022): use polynomial sieves: $\mathbb{E}\left[\left(\sigma_j^{-1}\left(s_t; \mathbf{x}_t^{(2)}, F\right) - X_t^{(1)}\beta^{(1)}\right)\phi_k\left(Z_{jt}\right)\right] = 0$
- Lu, Shi, Tao (JE 2023): use partially linear model: $\log\left(s_{jt}/s_{0t}\right) = X'_{1,jt}\beta^0 + \psi^0\left(X_{2,jt}; S_{J,t}\right) + \xi_{jt}$ where $\psi^0\left(x_{2,jt}; IV_{J,t}\right) = \log\left[\frac{\int \frac{\exp\left(x'_{2,jt}v\right)}{\exp\left(IV_{J,t}(v)\right)} f^0(v)\, dv}{\int \frac{1}{\exp\left(IV_{J,t}(v)\right)} f^0(v)\, dv}\right]$.

But these are still only as good as characteristics.

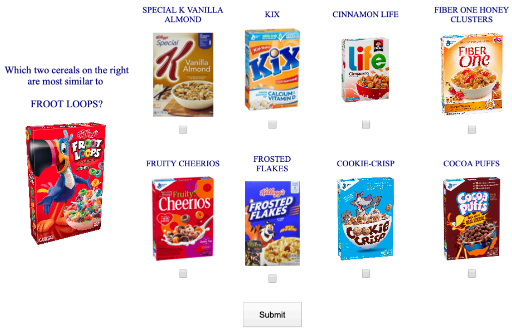
Low Rank Matrix Factorization: aka Netflix Prize



- Even if Ratings are sparse, we fit the observed cells and predict the rest!
- Idea: Approximate with a low rank (M) factor model.

Unobserved Characteristics: Magnolfi Maclure Sorensen (2023)

FIGURE 1: Sample survey page



What if we could first estimate unobserved characteristics?

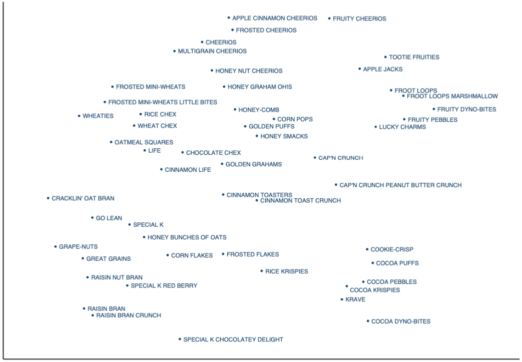
- Is j more similar to k or l?
- Use embedding procedure to calculate what amounts to a likelihood

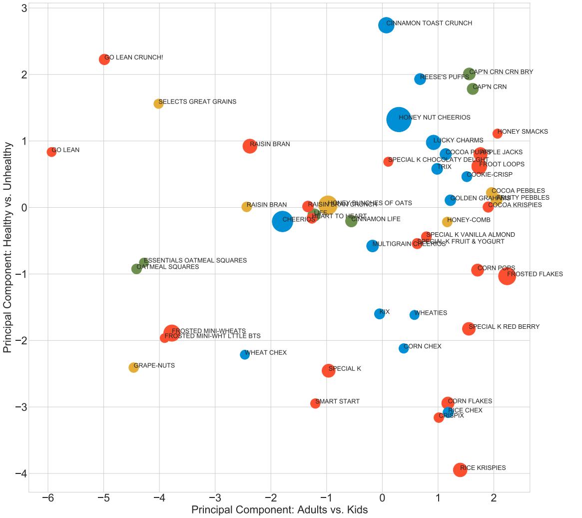
$$\max_{\mathbf{X}\in\mathbb{R}^{m\times J}} \ln\left(\frac{f(||x_l - x_j||, \alpha)}{f(||x_l - x_j||, \alpha) + f(||x_k - x_j||, \alpha)}\right)$$

- Get a $m \times J$ matrix with m factors (embeddings).
- Idea: m is small (like 3-4).

Unobserved Characteristics: Magnolfi Maclure Sorensen (2023)

FIGURE 2: Plot of two-dimensional embedding





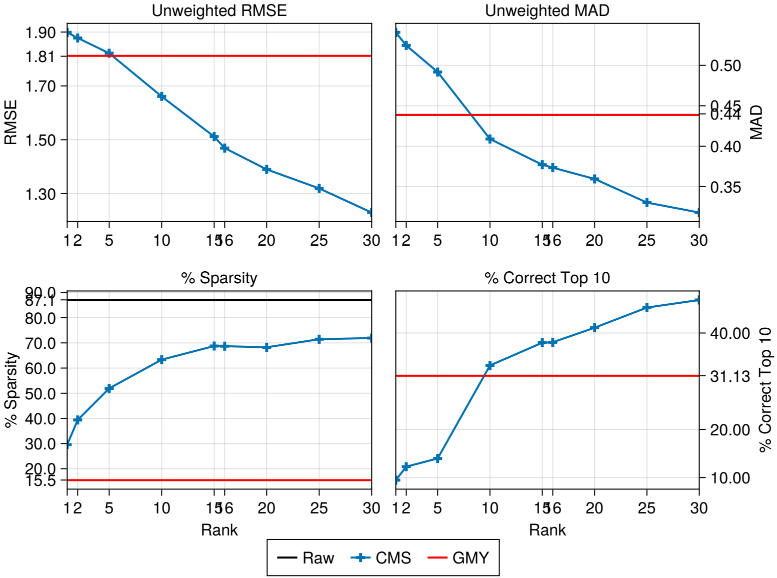
Conlon, Mortimer, Sarkis (2024): Structural Low Rank Approximations

Fix $\operatorname{rank}(\mathbf{D}(\mathbf{S},\pi)) = I$, and for each choice of I solve:

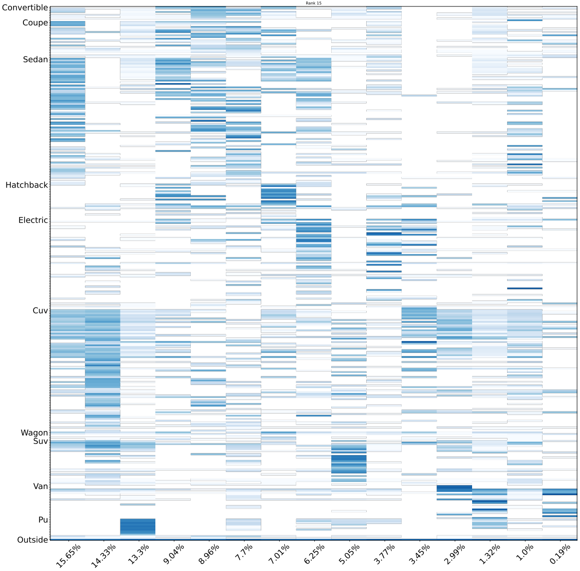
$$\min_{(\mathbf{S},\pi)\geqslant 0} \left\|\mathscr{P}_{\Omega}(\mathscr{D} - \mathbf{D}(\mathbf{S},\pi))\right\|_{\ell_2} + \lambda \left\|\mathcal{S} - \mathbf{S}\,\pi\right\|_{\ell_2} \text{ with } \left\|\pi\right\|_{\ell_1} \leqslant 1, \quad \left\|\mathbf{s_i}\right\|_{\ell_1} \leqslant 1.$$

- Goal: estimate $\mathbf{s_i}$ (choice probabilities) and corresponding weights π_i (Finite Mixture) in product space
  - Consistent with $U_{ij} = V_{ij} + \varepsilon_{ij}$ and logit error.
- Constraints: Choice probabilities s_{ij} sum to one, type weights π_i sum to one.
  - ℓ_1 constraints lead to sparsity.
- Idea: Control the rank by limiting I directly
  - Use cross validation to select # of types I and Lagrange multiplier λ.
- Matrix completion: We can construct estimates of $\mathbf{D}(\mathbf{S},\pi)$ including elements of $\mathscr{P}_{\overline{\Omega}}$.

In-Sample Performance



Profiles of Types (Rank 15)



Top Substitutes: Ford F-Series

Model	Raw	Logit	CMS I=15	CMS I=30	GMY
Ram Pickup	24.59	0.88	21.46	22.23	19.4
Gmc Sierra	20.29	0.61	14.97	21.92	17.27
Chevrolet Silverado	15.62	0.78	13.408	19.63	33.62
Toyota Tundra	12.98	0.55	16.32	12.79	2.29
Toyota Tacoma	6.31	0.76	3.39	3.13	2.83
Chevrolet Colorado	4.64	0.63	3.22	2.86	2.87
Gmc Canyon	2.3	0.3	0.76	1.38	1.02
Nissan Frontier	1.63	0.43	0.92	1.69	0.61
Jeep Wrangler	1.59	0.69	1.33	0.94	0.06
Nissan Titan	0.7	0.05	1.18	1.17	0.18
Ford Explorer	0.63	0.38	0.16	0.14	0.71

Thanks!

Reach me at cconlon@stern.nyu.edu

Slides at `bit.ly/conlon_IIOC`

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.

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., & 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.

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.