What's new in Micro Data?

Chris Conlon

Fall 2025

NYU Stern

What’s new with micro data?

Micro Data: Motivation

Data are a lot better than they were in 1995. It is common to observe:

- Choices of individuals, demographic characteristics, etc.
  - Examples: Nielsen Panelist Data, Kantar Worldpanel, IRI, etc.
- Higher frequency data, now $M_t \to \infty$ may not be a good assumption.
  - Not every product gets purchased every hour or every day.
  - Examples: Taxis/Uber, Daily airline bookings, hotel stays, AirBnB, Amazon, etc.

How can we make the best use of the data that we see?

Common Approaches

1. Use control function approach of Petrin and Train (2010)
2. Fully specifiy likelihood with a parametric assumption on $f(\xi)$ (Jiang Machanda Rossi (2009))
3. Add "micro-moments" a la Petrin (2001) or Micro BLP (2004) but otherwise aggregate data.
   - How do we choose them? Are we throwing away information?
4. Add aggregate data moments to maximum likelihood (Grieco, Murry, Pinkse, Sagl 2023)

$$
\begin{aligned}
u_{ijt} &= \beta_i\, x_{jt} - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt} \\
p_{jt} &= f(x_{jt}, z_{jt}) + e_{jt} \\
\xi_{jt} &= \lambda e_{jt} + \zeta_{jt} \text{ with } \mathbb{E}[\zeta_{jt} \mid x_{jt}, z_{jt}] = 0
\end{aligned}
$$

1. Run a first-stage for price to get $\widehat{f}(\cdot)$ and $\hat{e}_{jt}$.
2. Plug in $\hat{e}_{jt}$ into inner IV regresion of BLP:

$$
\delta_{jt}(\theta_2) = x_{jt}\beta - \alpha p_{jt} + \lambda \hat{e}_{jt} + \zeta_{jt}
$$

3. We can ignore λ, goal is to “clean” ξ_{jt} of the endogeneity and get $\widehat{\alpha}$ correct.

Control Functions: What's the point?

- We can use the same $\hat{e}_{jt}$ approach and just do MLE like McFadden and Train insead. lambda, goal is to "clean" ξ_{jt} of the endogeneity.
- This means we can use full data on individual choices, while also addressing endogeneity of prices.
- Possibly more efficient but with really strong assumptions of $f(\cdot)$ being correctly specified
- Controversy: We need a consistent estimate for $\hat{e}_{jt}$ so $f(x_{jt}, z_{jt})$ needs to be the "true" mapping from observables to endogenous prices.

Fully specified $f(\xi, \omega)$ (Jiang Machanda Rossi (2009))

These approaches come from the Bayesian Marketing literature and "Bayesian IV"

1. Specify a reduced form pricing equation:

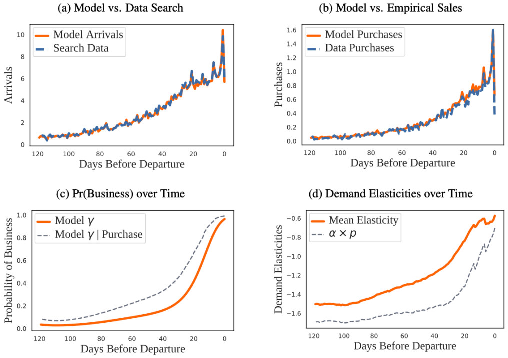
$$p_{jt} = \gamma z_{jt} + \omega_{jt}$$

2. Specify a parametric (joint) distribution for $f(\xi, \omega)$
   - Idea: be as flexible as possible allowing for correlation and mixtures of normals. (Bayesian Nonparametrics)
   - Challenge: working out the Jacobian from $s_{jt} \rightarrow \xi_{jt}$ (just as in BLP).
3. Advantage: now we can do individual data MLE. Can also allow for very flexible distributions of $f(\beta_i \mid \theta)$ if we do MCMC.

Main Complaint: Markups are a function of everything inclcuding $\omega_{-j,t}$ and $\xi_{-j,t}$ (maybe?)

Fully specified $f(\xi, \omega)$: Hortacsu, Natan, Parsley, Schweig, Williams (2023)

Figure 6: Demand Estimates



Note: The horizontal axis of all plots denotes the negative time index, e.g. zero corresponds to the last day before departure. (a) Normalized model fit of searches with data searches. (b) Model fit of product shares with empirical shares. (c) Fitted values of γ_t over time, along with the probability a consumer is business conditional on purchase. (d) Mean product elasticities over time, along with the least and most elastic flights.

A nice extension/application of this approach

- Consider high-frequency airline data where markets M_t are small
- Allow for poisson arrivals of consumers to the market who choose with mixed logit probabilities
- Not much individual data. Allow for business/leisure travelers (unobserved types)
- Mixture of Dirichlet Process for (ξ, ω)
- One really good z_{jt} (estimated shadow prices)

Combined Likelihood Approaches (GMPS 2023)

- The researcher observes market-level data of purchases, and can construct market shares

$$s_{jm} = \frac{1}{N_m} \sum_{i=1}^{N_m} d_{ijm} \tag{1}$$

  where N_m is the market size.
- For a subset of S_m consumers (denoted by the dummy D_{im}) the researcher observes $\{(d_{i\cdot m}, y_{im})\}$

Previous Approach: two-step estimator

- Estimate $(\widehat{\delta}, \widehat{\theta})$ by maximizing the likelihood

$$\mathcal{L}(\delta, \theta) = \sum_i \sum_j \sum_m d_{ijm} \log \int \underbrace{\frac{\exp\left(\delta_{jm} + \mu^y_{ijm} + \mu^\nu_{ijm}\right)}{\sum_{l=0}^{Jm} \exp\left(\delta_{jm} + \mu^y_{ijm} + \mu^\nu_{ijm}\right)}}_{s_{jm}(y_{im}, \nu; \theta, \delta)} dF(\nu)$$

and recover $\widehat{\beta}$ by running IV regression on $\widehat{\delta}$.

Bayer, Ferreira, and McMillan (2007) Bayer and Timmins (2007)

- Uses only within-market variation to estimate θ, which can break identification.

- GMPS 2023 propose the following estimator:

$$(\hat{\beta}, \hat{\theta}, \hat{\delta}) = \arg\min_{\beta,\theta,\delta} \underbrace{\left(-\log \hat{L}(\theta, \delta) + \hat{\Pi}(\beta, \delta) \right)}_{\hat{\Omega}(\beta,\theta,\delta)} \tag{2}$$

where

1. the **MDLE**

$$\log \hat{L}(\theta, \delta) = \sum_{m=1}^{M} \sum_{j=0}^{J_m} \sum_{i=1}^{N_m} d_{ijm} \left(D_{im} \log \pi_{jm}^{y_{im}}(\theta, \delta) + (1 - D_{im}) \log \pi_{jm}(\theta, \delta) \right) \tag{3}$$

$$= \sum_{m=1}^{M} \sum_{j=0}^{J_m} \sum_{i=1}^{N_m} D_{im} d_{ijm} \log \frac{\pi_{jm}^{y_{im}}}{\pi_{jm}} + \sum_{m=1}^{M} N_m \sum_{j=0}^{J_m} s_{jm} \log \pi_{jm} \tag{4}$$

If $N_m = S_m$ this estimator $\leftrightarrow$ mixed-logit estimator in previous slides.
If $S_m = 0$ maximizing the likelihood $\leftrightarrow$ imposing share constraints

2. the **product level moments**

$$\hat{\Pi}(\beta,\delta) = \frac{1}{2}\hat{m}^\top(\beta,\delta)\hat{\mathcal{W}}\hat{m}^\top(\beta,\delta) \tag{5}$$

where $\hat{m}^\top(\beta,\delta) = \sum_{m=1}^{M}\sum_{j=1}^{J_m} b_{jm}(\delta_{jm} - \beta^\top x_{jm})$.
If $d_b = d_\beta$ (exact identification) $\leftrightarrow$ two-step estimator
If $d_b > d_\beta$ (overidentification): both terms contribute to the estimation of θ, δ

- Show asymptotic equivalence of MDPLE with the GMM estimator

$$(\hat{\beta}, \hat{\theta}, \hat{\delta}) = \arg\min_{\beta,\theta,\delta} \frac{1}{2} \begin{bmatrix} \hat{m}^\top & \partial_{\psi^\top} \log \hat{L} \end{bmatrix} \begin{bmatrix} \hat{\mathcal{W}} & 0 \\ 0 & \hat{\mathcal{W}}_L \end{bmatrix} \begin{bmatrix} \hat{m} \\ \partial_{\psi} \log \hat{L} \end{bmatrix} \tag{6}$$

- Show that the estimator above is efficient
- Trick: Also setup problem to be convex in δ

Properties: conformant convergence

- Conformant: convergence rates adjust depending on alternative divergence rates of $\{N_m\}$, S, J and variation in the data
- β is identified only from $\hat{\Pi}$: convergence rate is always $\sqrt{J}$.
- Converge rates for θ, δ depend on whether θ_z is fixed, and whether it is 0 or $\neq 0$.

	rate		contributing term(s)	
case	θ^z	θ^ν, δ	for θ^z	for θ^ν
$S/J \to \infty$, $\theta^z \neq 0$	$\sqrt{S}$	$\sqrt{S}$	$\log \hat{L}$	$\log \hat{L}$
$S/J \to \infty$, $\theta^z = 0$	$\sqrt{S}$	$\sqrt{J}$	$\log \hat{L}$	$\hat{\Pi}$
$S/J \to c$, $\theta^z \neq 0$	$\sqrt{J}$	$\sqrt{J}$	both	both
$S/J \to c$, $\theta^z = 0$	$\sqrt{J}$	$\sqrt{J}$	both	$\hat{\Pi}$
$S/J \to 0$	$\sqrt{J}$	$\sqrt{J}$	$\hat{\Pi}$	$\hat{\Pi}$

- Variation in the micro data alone is sufficient to identify θ_z, θ^ν and δ if the micro sample is large enough and demographic variation affects choice probabilities substantially.

Limitations of the MDPLE estimator

- Incorporating supply-side moment conditions, or other moments that may make the MDPLE objective function non-convex.
- Limited gains compared to other approaches (in particular, the Optimal Micro BLP Estimator) when $\frac{S_m}{N_m} \to 0$ as $N \to \infty$.
- Assumes fully-compatible dataset (not true, e.g., if the researcher has censored data).

Micro BLP Approaches

Micro BLP is used a lot

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

- Many empirical IO papers use the “micro BLP” approach
  1. Impose the Berry, Levinsohn, and Pakes (1995) share constraint (unlike Grieco et al. 2023)
  2. Stack product-level or “aggregated” moments with “micro” moments from consumer surveys

Towards a standardized framework

- Despite the popularity of micro BLP, there's no standardized framework
  - Most papers use different notation to incorporate micro data in problem-specific ways
- So econometric work is limited, with a few exceptions
  - Berry and Haile (2014, 2022): Nonparametric identification, including with micro data
  - Myojo and Kanazawa (2012): Extend Berry, Linton, and Pakes (2004) asymptotics to Petrin (2002)
- We extend our BLP "best practices" (Conlon & Gortmaker, 2020) to the case with micro data
  1. Provide a standardized framework that covers most cases
  2. Derive practical advice from econometrics, simulations, examples
  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 $w_{d i_n j_n t_n}$

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

- *"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)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[y_i \mid j \in \mathcal{J}]$	Barwick et al. (2017)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$
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)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$	Wollmann (2018)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Goeree (2008)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$	S. Li (2018)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$
Ciliberto and Kuminoff (2010)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$	Y. Li et al. (2018)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$
Nakamura and Zerom (2010)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$	Backus et al. (2021)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$, $\mathbb{C}(x_j, y_i \mid j \neq 0)$
Beresteanu and Li (2011)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$	Grieco et al. (2021)	$\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$, $\mathbb{C}(x_j, x_{k(\text{-}j)} \mid j, k \neq 0)$
S. Li (2012)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[y_i \mid j \in \mathcal{J}]$	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 and Pinter (2022)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Starc (2014)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$	Döpper et al. (2022)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$
Ching et al. (2015)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$	Bodéré (2023)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$, $\mathbb{E}[x_j \mid i \in \mathcal{I}, j \neq 0]$
S. Li et al. (2015)	$\Pr(j \in \mathcal{J} \mid i \in \mathcal{I})$	Montag (2023)	$\mathbb{C}(x_j, y_i \mid j \neq 0)$, $\mathbb{C}(x_j, x_{k(\text{-}j)} \mid j, k \neq 0)$
Nurski and Verboven (2016)	$\mathbb{E}[y_i \mid j \in \mathcal{J}]$, $\mathbb{C}(x_j, y_i \mid j \neq 0)$	Conlon and 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!

Optimal micro moments

- What in the full micro data $\{(t_n, j_n, y_{i_n t_n})\}_{n \in \mathcal{N}_d}$ is most informative about θ_0? The score!

$$v_{ijt}(\theta_0) = \frac{\partial \log \Pr(t_n = t, j_n = j, y_{i_n t_n} = y_{it} \mid n \in \mathcal{N}_d)}{\partial \theta_0'}$$

- Inspecting score expressions gives intuition for which micro moments perform well

$$\underbrace{v_{ijt}(\Pi_0) \approx x_{jt} y_{it}}_{\text{Similar to“}\mathbb{C}(x_{jt}, y_{it} \mid j \neq 0)\text{”}} \qquad \underbrace{v_{ijkt}(\Sigma_0) \approx x_{jt} + x_{k(\text{-}j)t}}_{\text{Similar to “}\mathbb{C}(x_{jt}, x_{k(\text{-}j)t} \mid j, k \neq 0)\text{”}} \qquad \underbrace{v_{ijt}(\beta_0) = 0}_{\text{No direct info}}$$

- Feasible to match $v_{ijt}(\hat{\theta})$ at consistent $\hat{\theta}$ in second GMM step (with optimal weights/IVs)
  - Asymptotically efficient among all share-constrained micro BLP estimators
  - Computationally efficient too, only need to compute scores once
  - Must observe and be willing to use all info in the full micro data!

Some words of caution

- 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
- 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 "$\mathbb{C}(x_{jt}, y_{it} \mid j \neq 0)$"

Micro BLP estimator

$$\hat{\theta} = \arg\min_{\theta} \hat{g}(\theta)' \widehat{W} \hat{g}(\theta), \quad \hat{g}(\theta) = \begin{bmatrix} \hat{g}_{\mathsf{A}}(\theta) \\ \hat{g}_{\mathsf{M}}(\theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{N_{\mathsf{A}}} \sum_t \sum_j (\hat{\delta}_{jt}(\Pi, \Sigma) - x_{jt}'\beta) z_{jt} \\ f_1(\overline{v}) - f_1(v(\Pi, \Sigma)) \\ \vdots \\ f_{M_{\mathsf{M}}}(\overline{v}) - f_{M_{\mathsf{M}}}(v(\Pi, \Sigma)) \end{bmatrix}$$

- Berry, Levinsohn, and Pakes's (1995) share constraint gives mean utilities $\hat{\delta}_{jt}(\Pi, \Sigma)$

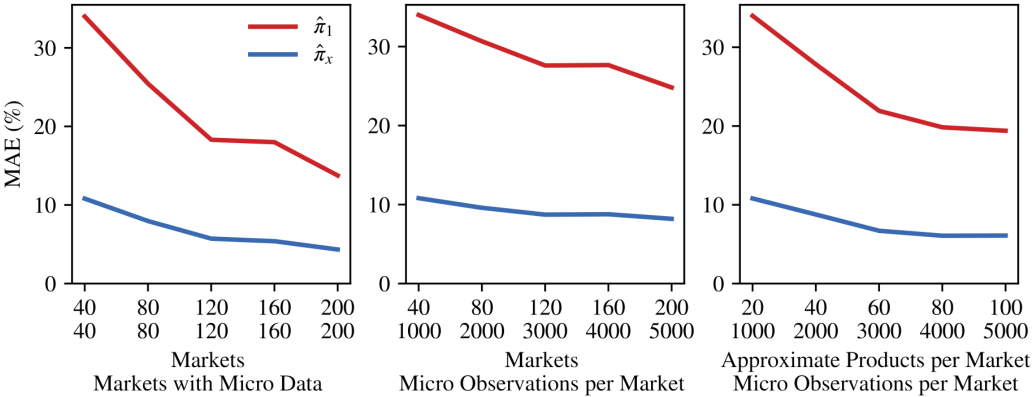
$$s_{jt} = \sum_i w_{it} \frac{\exp[\hat{\delta}_{jt}(\Pi, \Sigma) + x_{jt}'(\Pi y_{it} + \Sigma \nu_{it})]}{1 + \sum_k \exp[\hat{\delta}_{kt}(\Pi, \Sigma) + x_{kt}'(\Pi y_{it} + \Sigma \nu_{it})]}$$

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

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

- Highlight this and other practical advice in Monte Carlo experiments
- Study 3 types of asymptotics, punchline is that desirable properties translate to finite samples

- Highlight this and other practical advice in Monte Carlo experiments
- Study 3 types of asymptotics, punchline is that desirable properties translate to finite samples



Collecting second choices for an empirical example

- We demonstrate how all this works with Nielsen data
  - Estimate pre-2017 demand for soft drinks in Seattle
- Counterfactual highlights diversion to the outside good
  - Predict effects of 2018 tax, compare with what happened (−22%)
- So also show how to run cheap second choice survey
  - Diversion ratios discipline counterfactual in interpretable way

	Scanner Data	& Household	& Diversion
Taxed Volume Change (%)	−30.1	−30.0	−16.5
	(1.4)	(1.5)	(1.7)
High − Low Income (pp)		2.0	1.0
		(0.8)	(0.9)

Collecting second choices for an empirical example

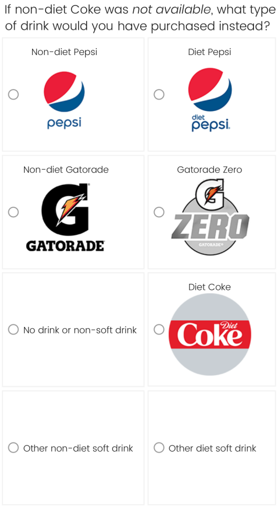
- We demonstrate how all this works with Nielsen data
  - Estimate pre-2017 demand for soft drinks in Seattle
- Counterfactual highlights diversion to the outside good
  - Predict effects of 2018 tax, compare with what happened (—22%)
- So also show how to run cheap second choice survey
  - Diversion ratios discipline counterfactual in interpretable way

	Scanner Data	& Household	& Diversion
Taxed Volume Change (%)	—30.1	—30.0	—16.5
	(1.4)	(1.5)	(1.7)
High — Low Income (pp)		2.0	1.0
		(0.8)	(0.9)

Collecting second choices for an empirical example

- We demonstrate how all this works with Nielsen data
  - Estimate pre-2017 demand for soft drinks in Seattle
- Counterfactual highlights diversion to the outside good
  - Predict effects of 2018 tax, compare with what happened (−22%)
- So also show how to run cheap second choice survey
  - Diversion ratios discipline counterfactual in interpretable way

	Scanner Data	& Household	& Diversion
Taxed Volume Change (%)	−30.1	−30.0	−16.5
	(1.4)	(1.5)	(1.7)
High − Low Income (pp)		2.0	1.0
		(0.8)	(0.9)



Replicating Petrin (2002) with PyBLP

```python
import numpy as np
import pandas as pd
from pyblp import data, Problem, Formulation, MicroDataset, MicroPart, MicroMoment, Optimization, Iteration

# Configure the aggregate problem: linear demand ("X1"), nonlinear demand ("X2"), marginal costs ("X3"), and demographics
problem = Problem(
    product_formulations=[
        Formulation('1 + hpwt + space + air + mpd + fwd + mi + sw + su + pv + pgnp + trend + trend2'),
        Formulation('1 + I(-prices) + hpwt + space + air + mpd + fwd + mi + sw + su + pv'),
        Formulation('1 + log(hpwt) + log(wt) + log(mpg) + air + fwd + trend * (jp + eu) + log(q)'),
    ],
    costs_type='log',
    agent_formulation=Formulation('1 + I(low / income) + I(mid / income) + I(high / income) + I(log(fs) * fv) + age + fs + mid + high'),
    product_data=pd.read_csv(data.PETRIN_PRODUCTS_LOCATION),
    agent_data=pd.read_csv(data.PETRIN_AGENTS_LOCATION),
)

# Configure the micro dataset: name, number of observations, and a function that computes sampling weights
dataset = MicroDataset("CEX", 29125, lambda t, p, a: np.ones((a.size, 1 + p.size)))

# Configure micro moment parts: names, datasets, and functions that compute micro values
age_mi_part      = MicroPart("E[age_i * mi_j]",      dataset, lambda t, p, a: np.outer(a.demographics[:, 5], np.r_[0, p.X2[:,  7]]))
age_sw_part      = MicroPart("E[age_i * sw_j]",      dataset, lambda t, p, a: np.outer(a.demographics[:, 5], np.r_[0, p.X2[:,  8]]))
age_su_part      = MicroPart("E[age_i * su_j]",      dataset, lambda t, p, a: np.outer(a.demographics[:, 5], np.r_[0, p.X2[:,  9]]))
age_pv_part      = MicroPart("E[age_i * pv_j]",      dataset, lambda t, p, a: np.outer(a.demographics[:, 5], np.r_[0, p.X2[:, 10]]))
fs_mi_part       = MicroPart("E[fs_i * mi_j]",       dataset, lambda t, p, a: np.outer(a.demographics[:, 6], np.r_[0, p.X2[:,  7]]))
fs_sw_part       = MicroPart("E[fs_i * sw_j]",       dataset, lambda t, p, a: np.outer(a.demographics[:, 6], np.r_[0, p.X2[:,  8]]))
fs_su_part       = MicroPart("E[fs_i * su_j]",       dataset, lambda t, p, a: np.outer(a.demographics[:, 6], np.r_[0, p.X2[:,  9]]))
fs_pv_part       = MicroPart("E[fs_i * pv_j]",       dataset, lambda t, p, a: np.outer(a.demographics[:, 6], np.r_[0, p.X2[:, 10]]))
inside_mid_part  = MicroPart("E[1{j > 0} * mid_i]",  dataset, lambda t, p, a: np.outer(a.demographics[:, 7], np.r_[0, p.X2[:,  0]]))
inside_high_part = MicroPart("E[1{j > 0} * high_i]", dataset, lambda t, p, a: np.outer(a.demographics[:, 8], np.r_[0, p.X2[:,  0]]))
mi_part          = MicroPart("E[mi_j]",              dataset, lambda t, p, a: np.outer(a.demographics[:, 0], np.r_[0, p.X2[:,  7]]))
sw_part          = MicroPart("E[sw_j]",              dataset, lambda t, p, a: np.outer(a.demographics[:, 0], np.r_[0, p.X2[:,  8]]))
su_part          = MicroPart("E[su_j]",              dataset, lambda t, p, a: np.outer(a.demographics[:, 0], np.r_[0, p.X2[:,  9]]))
pv_part          = MicroPart("E[pv_j]",              dataset, lambda t, p, a: np.outer(a.demographics[:, 0], np.r_[0, p.X2[:, 10]]))
mid_part         = MicroPart("E[mid_i]",             dataset, lambda t, p, a: np.outer(a.demographics[:, 7], np.r_[1, p.X2[:,  0]]))
high_part        = MicroPart("E[high_i]",            dataset, lambda t, p, a: np.outer(a.demographics[:, 8], np.r_[1, p.X2[:,  0]]))
```

```python
# Configure micro moments: names, observed values, parts, and functions that combine parts
ratio = lambda v: v[0] / v[1]
gradient = lambda v: [1 / v[1], -v[0] / v[1]**2]
micro_moments = [
    MicroMoment("E[age_i | mi_j]",      0.783,  [age_mi_part,      mi_part],   ratio, gradient),
    MicroMoment("E[age_i | sw_j]",      0.730,  [age_sw_part,      sw_part],   ratio, gradient),
    MicroMoment("E[age_i | su_j]",      0.740,  [age_su_part,      su_part],   ratio, gradient),
    MicroMoment("E[age_i | pv_j]",      0.652,  [age_pv_part,      pv_part],   ratio, gradient),
    MicroMoment("E[fs_i | mi_j]",       3.86,   [fs_mi_part,       mi_part],   ratio, gradient),
    MicroMoment("E[fs_i | sw_j]",       3.17,   [fs_sw_part,       sw_part],   ratio, gradient),
    MicroMoment("E[fs_i | su_j]",       2.97,   [fs_su_part,       su_part],   ratio, gradient),
    MicroMoment("E[fs_i | pv_j]",       3.47,   [fs_pv_part,       pv_part],   ratio, gradient),
    MicroMoment("E[1{j > 0} | mid_i]",  0.0794, [inside_mid_part,  mid_part],  ratio, gradient),
    MicroMoment("E[1{j > 0} | high_i]", 0.1581, [inside_high_part, high_part], ratio, gradient),
]
# Configure two-step minimum distance: starting values, optimization, and micro moments
problem_results = problem.solve(
    sigma=np.diag([3.23, 0, 4.43, 0.46, 0.01, 2.58, 4.42, 0, 0, 0, 0]),
    pi=np.array([
        [0, 0,     0,     0,    0,    0, 0, 0, 0],
        [0, 7.52, 31.13, 34.49, 0,    0, 0, 0, 0],
        [0, 0,     0,     0,    0,    0, 0, 0, 0],
        [0, 0,     0,     0,    0,    0, 0, 0, 0],
        [0, 0,     0,     0,    0,    0, 0, 0, 0],
        [0, 0,     0,     0,    0,    0, 0, 0, 0],
        [0, 0,     0,     0,    0,    0, 0, 0, 0],
        [0, 0,     0,     0,    0.57, 0, 0, 0, 0],
        [0, 0,     0,     0,    0.28, 0, 0, 0, 0],
        [0, 0,     0,     0,    0.31, 0, 0, 0, 0],
        [0, 0,     0,     0,    0.42, 0, 0, 0, 0],
    ]),
    optimization=Optimization('bfgs', {'gtol': 1e-4}),
    iteration=Iteration('squarem', {'atol': 1e-13}),
    se_type='clustered',
    W_type='clustered',
    micro_moments=micro_moments,
)
```

- We've tried our best to make all this easy to do with PyBLP, including optimal micro moments
    - E.g. can estimate Petrin's (2002) model with $<$ 100 lines of code

- If you have relatively complete data on individual decisions, do MDLE
- If you have mostly aggregate data and some survey moments do microBLP.
- If you have very high-frequency data with lots of zeros, you are probably stuck doing some full likelihood Bayesian approach.

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., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, *82*(5), 1749–1797.

Berry, S. T., & Haile, P. A. (2022). Nonparametric identification of differentiated products demand using micro data.

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.

Berry, S. T., Linton, O. B., & Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, *71*(3), 613–654.

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., & Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. *RAND Journal of Economics*, *51*(4), 1108–1161.

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.

Myojo, S., & Kanazawa, Y. (2012). On asymptotic properties of the parameters of differentiated product demand and supply systems when demographically categorized purchasing pattern data are available. *International Economic Review*, *53*(3), 887–938.

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.