Efficient Estimation of
Random Coefficients Demand Models using Product and Consumer Datasets

Paul L. E. Grieco — Penn State
Charles Murry — Boston College
Joris Pinkse — Penn State
Stephan Sagl — Penn State

Why another (micro) BLP methods paper?

- **Applicability:**
  "nests" mixed logit, micro/macroBLP datasets & assumptions.
- **Efficiency:**
  makes full use of observed data and assumptions.
- **Inference:**
  avoids strong assumptions; accounts for sampling in aggregate shares.

How is this accomplished

- Exploit likelihood of micro-data:
  no picking moments.
- Replace share constraint with "macro likelihood":
  allows micro-data to inform share estimates.
- Proper weighting between datasets.
- Introduce product-moments via penalty:
  adaptive rate of convergence.
- No computational penalty (objective function profiles).

The Model

A consumer i in market m with observable characteristics z_{im} purchases product j with probability:

$$\pi_{jm}^{z_{im}}(\psi) = \mathbb{P}(y_{ijm} = 1 | z_{im}; \psi) = \int \underbrace{\frac{\exp(\delta_{jm} + \mu_{ijm}^{z} + \mu_{ijm}^{\nu})}{\sum_{g=0}^{J_m} \exp(\delta_{gm} + \mu_{igm}^{z} + \mu_{igm}^{\nu})}}_{s_{jm}(z_{im}, \nu; \psi)} \, dF_m(\nu),$$

where δ_{jm} absorbs product-market quality shock:

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

and ψ collects δ and parameters of μ.

Intuition from "missing" demographic data:

Suppose we observed y_{ijm} for all consumers but z_{im} only if $D_{im} = 1$. The loglikelihood would be:

$$\mathsf{LL}(\psi) = \sum_{m=1}^{M} \sum_{j=0}^{J_m} \sum_{i=1}^{N_m} y_{ijm} \left(D_{im} \log \pi_{jm}^{z_{im}}(\psi) + (1 - D_{im}) \log \pi_{jm}^{D=0}(\psi) \right),$$

where we can compute,

$$\pi_{jm}^{D=0}(\psi) = \int \mathbb{P}(y_{ijm} = 1 \cap D_{im} = 0 \mid z_{im} = z)\, dG_m(z).$$

Our loglikelihood

We observe a consumer sample $\{y_{ijm}, z_{im}\}$ and product shares

$$s_{jm} = \frac{1}{N_m} \sum_{i=1}^{N_m} \mathbb{1}_{y_{ijm}=1},$$

so rewrite loglikelihood:

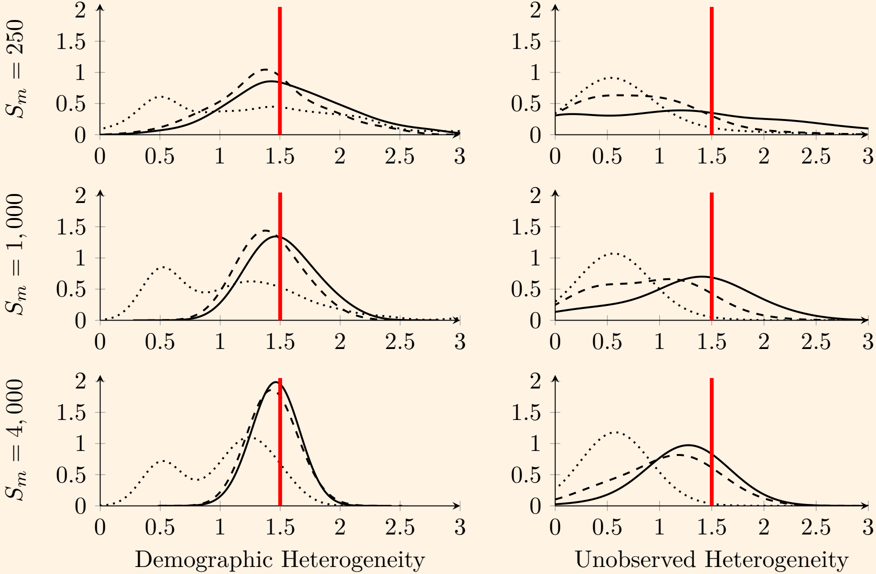
$$\mathrm{LL}(\psi) = \underbrace{\sum_{m=1}^{M} \sum_{j=0}^{J_m} \sum_{i=1}^{N_m} D_{im} y_{ijm} \log \frac{\pi_{jm}^{z_{im}}(\psi)}{\pi_{jm}^{D=1}(\psi)}}_{\text{micro}} + \underbrace{\sum_{m=1}^{M} N_m \sum_{j=0}^{J_m} s_{jm} \log \pi_{jm}(\psi)}_{\text{macro}},$$

recover $(\hat{\alpha}, \hat{\beta}) = \Xi\hat{\delta}$ with typical projection, $\Xi = (X^\intercal P_v X)^{-1} X^\intercal P_v$.

Contrast with GMM approach

- FOCs of macro term are the "δ-inversion" (Berry 1994) but they are not imposed as a constraint since micro term exists.
  - Small or heterogeneously sized markets.
  - Sampling of micro-data.
- Replace aggregated micro-moments with micro-likelihood.
- Straightforward inference that accounts for sampling error in macro shares—matters when shares are small!

GMM (dot) v. LL + Share Constraint (dash) v. MDLE (solid)



Are random coefficients (μ^{ν}_{ijm}) identified?

- If demographics shift utility ($\mu^{z}_{ijm} \neq 0$), yes! Just as in a mixed-logit.
- However, if $\mu^{z}_{ijm} \approx 0$, identification will be weak. How close is close? Depends on consumer sample size $S_m = \sum_i D_{im}$.
- Product level moments introduce additional assumptions to address non or weak identification.
  - Benefit: stronger exogeneity assumptions (powerful, if true).
  - Cost: Moments coverage at rate J (number of products), not S_m (number of sampled consumers).

Best of both worlds: Penalized Estimator

We add a penalty term to our estimator to incorporate product moments:

$$LL(\psi) - \frac{1}{J}\hat{m}^{\intercal}(\alpha, \beta, \delta)\hat{\mathcal{W}}\hat{m}(\alpha, \beta, \delta)$$

where,

$$\hat{m}(\alpha, \beta, \delta) = \sum_{m=1}^{M} \sum_{j=1}^{J_m} b_{jm}(\delta_{jm} - \beta^{\intercal} x_{jm} - \alpha p_{jm}),$$

and b_{jm} are instruments such that $E[b_{jm}\xi_{jm}] = 0$.

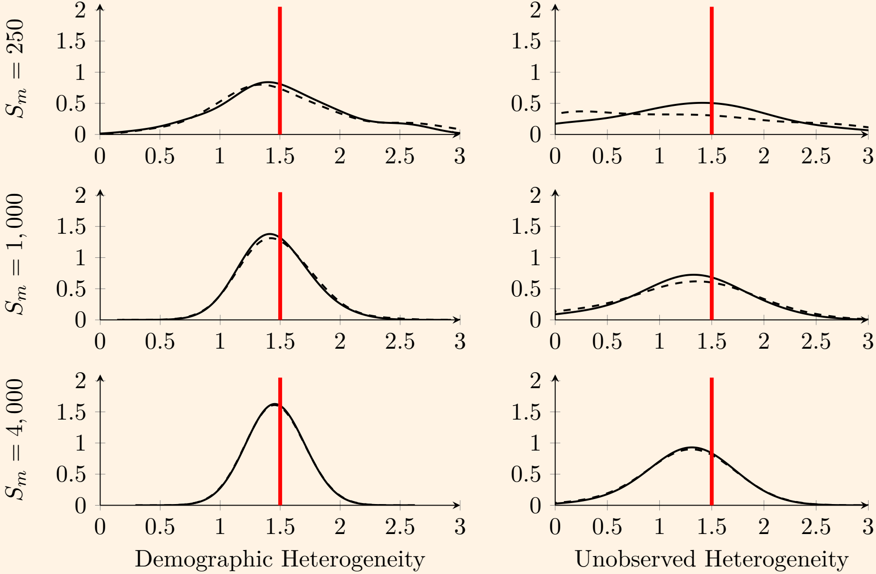
Note penalty also estimates (α, β), need sufficient dimension in b.

Adaptive Convergence

	$\sqrt{S/J} \times \mu^z$ **small**		$\sqrt{S/J} \times \mu^z$ **large**	
Estimation method	$\hat{\beta}$	$\hat{\theta}^\nu$	$\hat{\beta}$	$\hat{\theta}^\nu$
Penalized loglikelihood	$\sqrt{J}$	$\sqrt{J}$	$\sqrt{J}$	$\sqrt{S}$
LL (no penalty)	weak identification		$\sqrt{J}$	$\sqrt{S}$
GMM: $\partial\hat{\Omega}$ and $\hat{\Pi}$	$\sqrt{J}$	$\sqrt{J}$	$\sqrt{J}$	$\sqrt{J}$

Table: Asymptotic rates of convergence with (strong and valid) product-level moments

MDLE (dashed) vs. Penalized-MDLE (solid)



Conclusion

- Efficient likelihood-based estimator for discrete choice demand.
- Incorporates
    - Consumer-level data
    - Market-level data
    - Product-level exogeneity restricitons
- Adaptive convergence rate for weak identification.
- Straightforward, correct inference on all model parameters and functionals (elasticities, diversion, counterfactuals).