Data-Driven Nests

Montreal IO Conference

June 26th, 2025

Milena Almagro
(Chicago Booth & NBER)

Kenneth Lai
(NYU)

Elena Manresa
(NYU)

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda and Weinstein (2006), Khandelwal (2010), Costinot, Donaldson, and Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda and Weinstein (2006), Khandelwal (2010), Costinot, Donaldson, and Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024), Couture et al. (2025)...

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda and Weinstein (2006), Khandelwal (2010), Costinot, Donaldson, and Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024), Couture et al. (2025)...
- Labor: Azar et al. (2022), Grigsby (2022)...

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda and Weinstein (2006), Khandelwal (2010), Costinot, Donaldson, and Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024), Couture et al. (2025)...
- Labor: Azar et al. (2022), Grigsby (2022)...
- Macro: Atkeson & Burnstein (2008), Edmond, Midrigan & Xu (2015), Berger, Herkenhoff & Mongey (2022)...

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda and Weinstein (2006), Khandelwal (2010), Costinot, Donaldson, and Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024), Couture et al. (2025)...
- Labor: Azar et al. (2022), Grigsby (2022)...
- Macro: Atkeson & Burnstein (2008), Edmond, Midrigan & Xu (2015), Berger, Herkenhoff & Mongey (2022)...

What's the right nesting structure?

Demand with nesting structures across products is ubiquitous in many fields:

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Björnerstedt and Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda and Weinstein (2006), Khandelwal (2010), Costinot, Donaldson, and Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024), Couture et al. (2025)...
- Labor: Azar et al. (2022), Grigsby (2022)...
- Macro: Atkeson & Burnstein (2008), Edmond, Midrigan & Xu (2015), Berger, Herkenhoff & Mongey (2022)...

In general, the practitioner faces the challenge of specifying the nesting structure ex-ante
$\longrightarrow$ Mis-specification of nests can lead to biased estimates Fosgerau et al. (2024)

This paper: Estimating nests in discrete choice models

- Methodology to estimate the nesting structure as well as preference parameters
- Nesting structure is recovered from **aggregate market share data + product characteristics**
- Two-step estimation procedure:
  1. Use k-means clustering to estimate the nesting structure Bonhomme and Manresa (2015):
     - Marginal effect of covariates is common across products in the same nest
  2. Estimate model parameters as if the groups where known Berry (1994)
- We exploit the structure of the model, the availability of many markets and of many products

Related literature

- **Discrete Choice Models of Random Utility with Nests:** McFadden (1978, 1981), Berry (1994), Verboven (1996), Cardell (1997), McFadden and Train (2000), Grigolon and Verboven (2014)
- **Group Fixed Effect Estimator:** Han & Moon (2010), Bonhomme & Manresa (2015), Choi and Okui (2024)
- **Alternative Grouping Structure:** Fosgerau, Monardo & De Palma (2024), Hortacsu, Lieber, Monardo & de Paula (ongoing)

Outline

Empirical model

Discrete choice model with nested logit shocks

- Consider the indirect utility model for agent i when choosing j in market m:

$$V_{ijm} = \delta_{jm} + \varepsilon_{ijm}$$

- Choice of j based on the maximization of the utiliy:

$$\mathbb{P}_{jm} = \mathbb{P}(V_{ijm} > V_{ij'm} \quad \forall j' \neq j)$$

Discrete choice model with nested logit shocks

- Consider the indirect utility model for agent i when choosing j in market m:

$$V_{ijm} = \delta_{jm} + \varepsilon_{ijm}$$

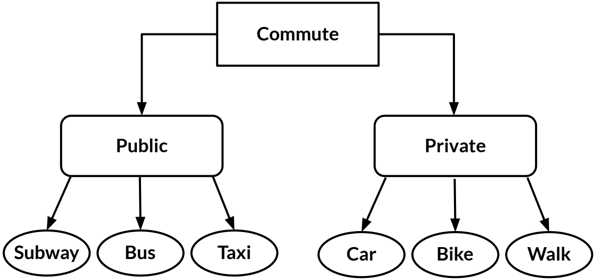
- Choice of j based on the maximization of the utiliy:

$$\mathbb{P}_{jm} = \mathbb{P}(V_{ijm} > V_{ij'm} \quad \forall j' \neq j)$$

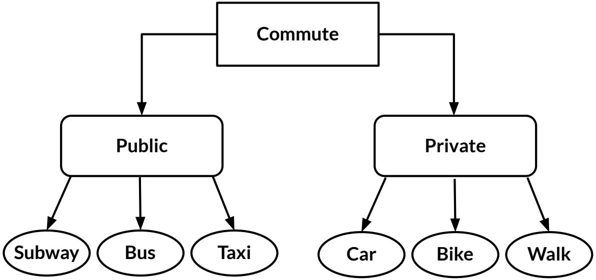
- Assume products are partitioned in K groups, and $(\varepsilon_{i1m}, \dots, \varepsilon_{iJm}) \sim \exp\Big(-\sum_{k=1}^{K} (\sum_{j \in B_K} e^{-\frac{\epsilon_j}{\sigma^{k(j)}}})^{\sigma^{k(j)}}\Big)$:

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \big(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \big)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \big(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^{l}}} \big)^{\sigma^{l}}}. \tag{1}$$

Nested logit as sequential choice



Nested logit as sequential choice



Choice of option j within nest $k(j)$

$$\mathbb{P}_{jm} = \underbrace{\left(\frac{\left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} \right)}_{\mathbb{P}_{k(j),m}} \underbrace{\left(\frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}}} \right)}_{\mathbb{P}_{j|k(j),m}}$$

Nested logit and substitution patterns

Group correlation allows for more flexible substitution patterns

$$\mathcal{E}^{j}_{j,'m} = \begin{cases} \frac{1}{\sigma^{k(j)}}\left(1-\sigma^{k(j)}\mathbb{P}_{jm}-(1-\sigma^{k(j)})\mathbb{P}_{jm|k(j)}\right)\frac{\partial\delta_{jm}}{\partial p_{jm}}p_{jm} & \text{if } j'=j \\ \left((\sigma^{k(j)}-1)\frac{1}{\sigma^{k(j)}}\mathbb{P}_{j'm|k(j)}-\mathbb{P}_{j'm}\right)\frac{\partial\delta_{j'm}}{\partial p_{j'm}}p_{j'm} & \text{if } j'\in B_{k(j)} \\ -\mathbb{P}_{j'm}\frac{\partial\delta_{j'm}}{\partial p_{j'm}}p_{j'm} & \text{if } j'\in B_{k'}\neq B_{k(j)}. \end{cases}$$

Nested logit and substitution patterns

Group correlation allows for more flexible substitution patterns

$$\mathcal{E}^{j}_{j,'m} = \begin{cases} \frac{1}{\sigma^{k(j)}}\left(1 - \sigma^{k(j)}\mathbb{P}_{jm} - (1-\sigma^{k(j)})\mathbb{P}_{jm|k(j)}\right)\frac{\partial \delta_{jm}}{\partial p_{jm}} p_{jm} & \text{if } j' = j \\ \left((\sigma^{k(j)} - 1)\frac{1}{\sigma^{k(j)}}\mathbb{P}_{j'm|k(j)} - \mathbb{P}_{j'm}\right)\frac{\partial \delta_{j'm}}{\partial p_{j'm}} p_{j'm} & \text{if } j' \in B_{k(j)} \\ -\mathbb{P}_{j'm}\frac{\partial \delta_{j'm}}{\partial p_{j'm}} p_{j'm} & \text{if } j' \in B_{k'} \neq B_{k(j)}. \end{cases}$$

σ^k can be interpreted as degree of within-group independence:

- Higher $\sigma^k \implies$ less within-group correlation $\implies$ less within-group substitution
- If $\sigma^k = 1 \implies$ back to logit
- $\sigma^k \in (0, 1)$ to be consistent with utility maximization

Outline

1. Empirical model
   1.1 Identification
   1.2 Estimation
2. Consistency and statistical properties
3. Monte Carlo
4. Extensions
   4.1 Choosing the number of groups
   4.2 Introducing Consumer's Heterogeneity
   4.3 Higher-level nesting structures
5. Empirical application: Demand for beer

Toward a linear regression equation

Recall:

$$\mathbb{P}_{jm} = \left(\frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}}} \right) \left(\frac{\left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} \right) = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} (\overbrace{\mathsf{IV}_m^k}^{\sum_{d \in B_k} e^{\frac{\delta_{dm}}{\sigma^k}}})^{\sigma^{k(j)}-1}}{\underbrace{\mathsf{IV}_m}_{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}}},$$

Taking logs:

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} (\mathsf{IV}_m^k)^{\sigma^{k(j)}-1}}{\mathsf{IV}_m} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma^{k(j)}} + (\sigma^{k(j)} - 1) \log \mathsf{IV}_m^{k(j)} - \log \mathsf{IV}_m,$$

Toward a linear regression equation

Recall:

$$\mathbb{P}_{jm} = \left(\frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}}} \right) \left(\frac{\left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} \right) = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \Big(\overbrace{\mathrm{IV}_m^k}^{\sum_{d \in B_k} e^{\frac{\delta_{dm}}{\sigma^k}}} \Big)^{\sigma^{k(j)}-1}}{\underbrace{\mathrm{IV}_m}_{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}}},$$

Taking logs:

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} (\mathrm{IV}_m^k)^{\sigma^{k(j)}-1}}{\mathrm{IV}_m} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma^{k(j)}} + (\sigma^{k(j)} - 1) \log \mathrm{IV}_m^{k(j)} - \log \mathrm{IV}_m,$$

Toward a linear regression equation

Assume $\delta_{jm} = \beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}$.

Toward a linear regression equation

Assume $\delta_{jm} = \beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}$.

Substituting:

$$\begin{aligned}\log \mathbb{P}_{jm} &= \frac{\beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}}{\sigma^{k(j)}} + (\sigma^{k(j)} - 1) \log \mathsf{IV}_m^{k(j)} - \log \mathsf{IV}_m \\ &= \underbrace{\beta_x^{k(j)}}_{\frac{\beta_x}{\sigma^k}} x_{jm} + \underbrace{\beta_p^{k(j)}}_{\frac{\beta_p}{\sigma^k}} p_{jm} + \underbrace{\lambda_m^{k(j)}}_{(\sigma^{k(j)}-1) \log \mathsf{IV}_m^{k(j)} - \log \mathsf{IV}_m} + \xi_{jm},\end{aligned}$$

with an slight abuse of notation for ξ_{jm}

Intuition of nest identification

$$\log \mathbb{P}_{jm} = \beta_x^{k(j)} x_{jm} + \beta_p^{k(j)} p_{jm} + \xi_{jm} + \lambda_m^{k(j)},$$

Intuition of nest identification

$$\log \mathbb{P}_{jm} = \beta_x^{k(j)} x_{jm} + \beta_p^{k(j)} p_{jm} + \xi_{jm} + \lambda_m^{k(j)},$$

- **Key observation:** Group-specific intercept and slope common for products in the same group!
- **Intuition:**
  - The marginal effect of variation in covariates varies by nest
  - IIA within a nest: only j and j' covariates affect $\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{j'm}}$
  - Not IIA across nests: If $k(j) \neq k(j')$, $\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{j'm}}$ also function of $\lambda_m^{k(j)}$ and $\lambda_m^{k(j')}$

Intuition of nest identification

$$\log \mathbb{P}_{jm} = \beta_x^{k(j)} x_{jm} + \beta_p^{k(j)} p_{jm} + \xi_{jm} + \lambda_m^{k(j)},$$

- **Key observation:** Group-specific intercept and slope common for products in the same group!
- **Intuition:**
  - The marginal effect of variation in covariates varies by nest
  - IIA within a nest: only j and j' covariates affect $\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{j'm}}$
  - Not IIA across nests: If $k(j) \neq k(j')$, $\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{j'm}}$ also function of $\lambda_m^{k(j)}$ and $\lambda_m^{k(j')}$

Endogenous prices

Assume z_{jm} exogenous shifter of prices.

Endogenous prices

Assume z_{jm} exogenous shifter of prices.

Pricing equation:

$$p_{jm} = \gamma_x x_{jm} + \gamma_z z_{jm} + v_{jm}, \quad \mathbb{E}[v_{jm}, \xi_{jm} | x_{jm}, z_{jm}] = 0$$

Substitute into:

$$\log \mathbb{P}_{jm} = \beta_x^{k(j)} x_{jm} + \beta_p^{k(j)} p_{jm} + \xi_{jm} + \lambda_m^{k(j)}$$

→ Classify on reduced-form:

$$\log \mathbb{P}_{jm} = \underbrace{\tilde{\beta}_x^{k(j)}}_{\beta_x^{k(j)} + \beta_p^{k(j)} \gamma_x} x_{jm} + \underbrace{\tilde{\beta}_z^{k(j)}}_{\beta_p^{k(j)} \gamma_z} z_{jm} + \underbrace{\nu_{jm}}_{\xi_{jm} + \beta_p^{k(j)} v_{jm}} + \lambda_m^{k(j)},$$

Outline

Estimation of groups

Group-fixed effect estimator defined by the following clustering problem:

$$\underset{\substack{k(1),\ldots,k(J)\\ \beta^1,\ldots,\beta^K,\lambda_1^1,\ldots,\lambda_M^K}}{\arg\min} \sum_{m=1}^{M} \sum_{j=1}^{J} \left(\log \mathbb{P}_{jm} - (\beta_z^k z_{jm} + \beta_x^k x_{jm} + \lambda_m^k) \right)^2$$

Combinatorial, non-convex problem!

Estimation of groups

Group-fixed effect estimator defined by the following clustering problem:

$$\underset{\substack{k(1),\dots,k(J) \\ \beta^1,\dots,\beta^K,\lambda_1^1,\dots,\lambda_M^K}}{\arg\min} \sum_{m=1}^{M} \sum_{j=1}^{J} \left(\log \mathbb{P}_{jm} - (\beta_z^k z_{jm} + \beta_x^k x_{jm} + \lambda_m^k) \right)^2$$

Combinatorial, non-convex problem!

Solution: two-step algorithm

1. Classify products using clustering algorithm following Bonhomme and Manresa (2015)
2. Conditional on classification, estimate preference parameters β and σ following Berry (1994)

Why two steps?

Two-step strategy
First Step: Classification

1. Let $(\beta^{1,0}, \dots, \beta^{K,0}, \lambda_1^{K,0}, \dots, \lambda_M^{K,0})$ be a starting value.

Two-step strategy

First Step: Classification

1. Let $(\beta^{1,0}, \dots, \beta^{K,0}, \lambda_1^{K,0}, \dots, \lambda_M^{K,0})$ be a starting value.
2. For $(\beta^{1,s}, \dots, \beta^{K,s}, \lambda_1^{K,s}, \dots, \lambda_M^{K,s})$, compute for all $j \in J$:

$$k(j)^{s+1} = \operatorname*{arg\,min}_{k \in \{1,\dots,K\}} \sum_{m=1}^{M} \Big(\log \mathbb{P}_{jm} - (\beta_z^{k,s} z_{jm} + \beta_x^{k,s} x_{jm} + \lambda_m^{k,s}) \Big)^2,$$

to compute grouping structure $\mathcal{B}^{s+1}$.

Two-step strategy

First Step: Classification

1. Let $(\beta^{1,0}, \dots, \beta^{K,0}, \lambda_1^{K,0}, \dots, \lambda_M^{K,0})$ be a starting value.
2. For $(\beta^{1,s}, \dots, \beta^{K,s}, \lambda_1^{K,s}, \dots, \lambda_M^{K,s})$, compute for all $j \in J$:

$$k(j)^{s+1} = \arg\min_{k \in \{1,\dots,K\}} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_z^{k,s} z_{jm} + \beta_x^{k,s} x_{jm} + \lambda_m^{k,s}) \right)^2,$$

   to compute grouping structure $\mathcal{B}^{s+1}$.
3. Compute:

$$(\beta^{1,s+1}, \dots, \beta^{K,s+1}, \lambda_1^{K,s+1}, \dots, \lambda_M^{K,s+1}) =$$
$$\arg\min_{\beta^1,\dots,\beta^K,\lambda_1^1,\dots,\lambda_M^K} \sum_{j=1}^{J} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_z^{k(j),s+1} z_{jm} + \beta_x^{k(j),s+1} x_{jm} + \lambda_m^{k(j),s+1}) \right)^2$$

Two-step strategy

First Step: Classification

1. Let $(\beta^{1,0}, \dots, \beta^{K,0}, \lambda_1^{K,0}, \dots, \lambda_M^{K,0})$ be a starting value.
2. For $(\beta^{1,s}, \dots, \beta^{K,s}, \lambda_1^{K,s}, \dots, \lambda_M^{K,s})$, compute for all $j \in J$:

$$k(j)^{s+1} = \underset{k\in\{1,\dots,K\}}{\arg\min} \sum_{m=1}^{M} \Big(\log \mathbb{P}_{jm} - (\beta_z^{k,s} z_{jm} + \beta_x^{k,s} x_{jm} + \lambda_m^{k,s}) \Big)^2,$$

   to compute grouping structure $\mathcal{B}^{s+1}$.

3. Compute:

$$(\beta^{1,s+1}, \dots, \beta^{K,s+1}, \lambda_1^{K,s+1}, \dots, \lambda_M^{K,s+1}) =$$

$$\underset{\beta^1,\dots,\beta^K,\lambda_1^1,\dots,\lambda_M^K}{\arg\min} \sum_{j=1}^{J} \sum_{m=1}^{M} \Big(\log \mathbb{P}_{jm} - (\beta_z^{k(j),s+1} z_{jm} + \beta_x^{k(j),s+1} x_{jm} + \lambda_m^{k(j),s+1}) \Big)^2$$

4. Repeat until convergence of parameters.

Two-step strategy

Second step estimation

Once we have grouping structure $\{k(1), ..., k(J)\}$, estimate β and σ as if groups were known

Two-step strategy

Second step estimation

Once we have grouping structure $\{k(1), \ldots, k(J)\}$, estimate β and σ as if groups were known

We follow Berry (1994):

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \beta_p p_{jm} + \beta_x x_{jm} + (\sigma^{k(j)} - 1) \log \mathbb{P}_{j,m|k(j)} + \xi_{jm}$$

Two-step strategy

Second step estimation

Once we have grouping structure $\{k(1), \ldots, k(J)\}$, estimate β and σ as if groups were known

We follow Berry (1994):

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \beta_p p_{jm} + \beta_x x_{jm} + (\sigma^{k(j)} - 1) \log \mathbb{P}_{j,m|k(j)} + \xi_{jm}$$

Instrument for prices, z_{jm} + instrument for $\log \mathbb{P}_{j,m|k(j)}$

Two-step strategy

Second step estimation

Once we have grouping structure $\{k(1), \ldots, k(J)\}$, estimate β and σ as if groups were known

We follow Berry (1994):

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \beta_p p_{jm} + \beta_x x_{jm} + (\sigma^{k(j)} - 1) \log \mathbb{P}_{j,m|k(j)} + \xi_{jm}$$

Instrument for prices, z_{jm} + instrument for $\log \mathbb{P}_{j,m|k(j)}$

$$\mathbb{P}_{j,m|k(j)} = \frac{e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma^{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma^{k(j)}}}}$$

Two-step strategy

Second step estimation

Once we have grouping structure $\{k(1), ..., k(J)\}$, estimate β and σ as if groups were known

We follow Berry (1994):

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \beta_p p_{jm} + \beta_x x_{jm} + (\sigma^{k(j)} - 1) \log \mathbb{P}_{j,m|k(j)} + \xi_{jm}$$

Instrument for prices, z_{jm} + instrument for $\log \mathbb{P}_{j,m|k(j)}$

$$\mathbb{P}_{j,m|k(j)} = \frac{e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma^{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma^{k(j)}}}} \implies w^x_{jm} = \frac{\exp(x_{jm})}{\sum_{j' \in k(j)} \exp(x_{j'm})}$$

Consistency and statistical properties

Regularity conditions with unknown nests

Regularity conditions to ensure consistent classification:

- K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$ Why $M \to \infty$?

Regularity conditions with unknown nests

Regularity conditions to ensure consistent classification:

- K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$ Why $M \to \infty$?
- Thin tails
- Variation in x_{jm}, z_{jm} conditional on groups
- $\lambda_m^{k(j)}$ is a random variable that diverges when $J \to \infty$. Conditions so they are well behaved:
  - Balanced nests: $|B_k| = O_p(J)$ for $k = 1, \ldots K$
  - Normalization by $\log J$ to stay in compact space + conditions so that $\log J\lambda_m^k(J)$ is well-defined
  - Sequence $(\sigma_{01,J}, \ldots, \sigma_{0K,J})_{J=1}^{\infty}$ such that $\lambda_m^{k(j)}$ is defined when J grows
- Group separation: cannot have $\beta^k = \beta^{k'}$ and $\lambda_m^k = \lambda_m^{k'}$ for all m for some $k \neq k'$

Regularity conditions with unknown nests

Regularity conditions to ensure consistent classification:

- K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$ Why $M \to \infty$?
- Thin tails
- Variation in x_{jm}, z_{jm} conditional on groups
- $\lambda_m^{k(j)}$ is a random variable that diverges when $J \to \infty$. Conditions so they are well behaved:
  - Balanced nests: $|B_k| = O_p(J)$ for $k = 1, \dots K$
  - Normalization by $\log J$ to stay in compact space + conditions so that $\log J\lambda_m^k(J)$ is well-defined
  - Sequence $(\sigma_{01,J}, \dots, \sigma_{0K,J})_{J=1}^{\infty}$ such that $\lambda_m^{k(j)}$ is defined when J grows
- Group separation: cannot have $\beta^k = \beta^{k'}$ and $\lambda_m^k = \lambda_m^{k'}$ for all m for some $k \neq k'$

Regularity conditions with unknown nests

Regularity conditions to ensure consistent classification:

- K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$ Why $M \to \infty$?
- Thin tails
- Variation in x_{jm}, z_{jm} conditional on groups
- $\lambda_m^{k(j)}$ is a random variable that diverges when $J \to \infty$. Conditions so they are well behaved:
  - Balanced nests: $|B_k| = O_p(J)$ for $k = 1, \dots K$
  - Normalization by $\log J$ to stay in compact space + conditions so that $\log J\lambda_m^k(J)$ is well-defined
  - Sequence $(\sigma_{01,J}, \dots, \sigma_{0K,J})_{J=1}^{\infty}$ such that $\lambda_m^{k(j)}$ is defined when J grows
- Group separation: cannot have $\beta^k = \beta^{k'}$ and $\lambda_m^k = \lambda_m^{k'}$ for all m for some $k \neq k'$

$\Rightarrow$ Asymptotic super-consistency to the infeasible estimator where the groups are known
$\rightarrow$ Convergence rate is $O_p(JM^{-\eta})$ for any $\eta > 0$

Monte Carlo

Monte Carlo design: Data generating process

Fix $K = 3$. Set $\sigma = (0.2, 0.3, 0.6)$. Classify products using uniform draws $k(j) \sim \mathcal{U}\{1, \ldots, K\}$

Models for the average utility δ_{jm} and prices p_{jm} are given by:

$$\delta_{jm} = \beta_p p_{jm} + \beta_x x_{jm} + \xi_{jm} \quad \text{and} \quad p_{jm} = mc_{jm} + \rho \cdot \xi_{jm}$$

Generate data as follows:

$$\begin{bmatrix} \mu_x^k \\ \mu_{mc}^k \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix} \right) \implies \begin{bmatrix} x_{jm} \\ mc_{jm} \\ \xi_{jm} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x^k \\ \mu_{mc}^k \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

Set $\rho = 0.3$

Results

Results of 50 Bootstrap iterations

J	M	Runtime	Matched	True	β_p -1	β_c 1	σ_1 0.2	σ_2 0.3	σ_3 0.6
100	10	00:02	0.996	Mean β	-0.992	0.991	0.189	0.297	0.602
				Std β	0.032	0.033	0.034	0.024	0.007
100	50	00:25	1.0	Mean β	-1.001	0.998	0.2	0.3	0.6
				Std β	0.01	0.011	0.001	0.002	0.003
100	100	01:07	1.0	Mean β	-1.0	1.0	0.2	0.3	0.6
				Std β	0.006	0.007	0.001	0.001	0.003
500	10	00:06	0.995	Mean β	-1.0	0.998	0.199	0.298	0.6
				Std β	0.015	0.02	0.006	0.016	0.003
500	50	07:14	1.0	Mean β	-1.0	0.999	0.2	0.3	0.6
				Std β	0.004	0.004	0.0	0.001	0.001
500	100	29:24	1.0	Mean β	-1.0	0.999	0.2	0.3	0.6
				Std β	0.003	0.003	0.0	0.0	0.001
1000	10	00:12	1.0	Mean β	-1.0	0.999	0.199	0.3	0.6
				Std β	0.007	0.007	0.003	0.001	0.002
1000	50	44:57	1.0	Mean β	-1.0	1.0	0.2	0.3	0.6
				Std β	0.002	0.003	0.0	0.0	0.001
1000	100	11:14	1.0	Mean β	-1.0	0.999	0.2	0.3	0.6
				Std β	0.002	0.002	0.0	0.0	0.001

Extensions

Outline

Choosing K: Cross-validation with Elbow method

So far we have assumed the number of groups is known.

In practice, we can also estimate the number of groups using a N**-fold cross-validation** procedure.

For $k \in \{1, \ldots, K\}$:

- Divide products into N equal subsets, $P_1, \ldots, P_N$.
- Pick subset P_n and estimate grouping structure and grouping parameters in the other $N-1$ parts.
- Classify products across estimated groups in part P_n and **compute out-of-sample MSE**

$$MSE_n(k) = \frac{1}{J \cdot M} \sum_{m=1}^{M} \sum_{j \in P_n} (y_j - \beta_{m,-n}^{k(j)} x_j - \lambda_{m,-n}^{k(j)})^2$$

- Take average across N folds:

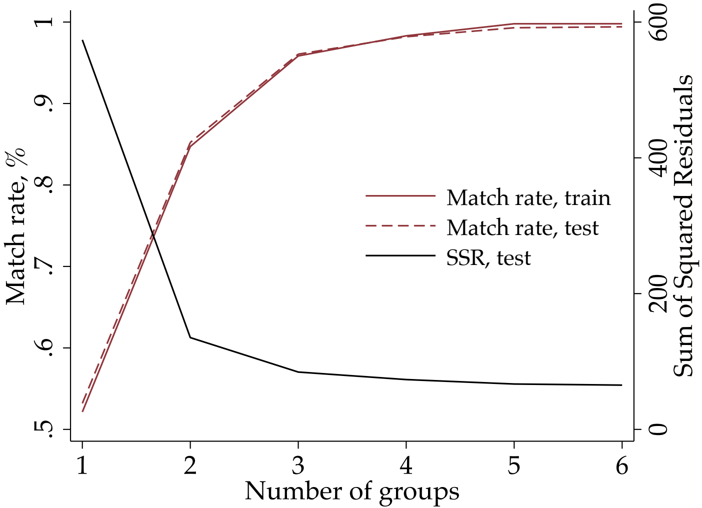
$$MSE(k) = \frac{1}{N} \sum_{n=1}^{N} MSE_n(k)$$

- Choose k according to

$$k^* = \{k(j) | \text{where slope of } MSE(k) \text{ changes}\}$$

Estimating the Number of Groups: Monte Carlo

Set $K^0 = 3$, the total number of products $J = 500$, and the total number of markets is $M = 10$.



Outline

Individual heterogeneity with observed conditional shares

Assume heterogeneity can be described by types t (e.g. income quintiles) and utility given by:

$$\delta^t_{jm} = \beta^t_p x_{jm} + \xi^t_{jm}$$

Individual heterogeneity with observed conditional shares

Assume heterogeneity can be described by types t (e.g. income quintiles) and utility given by:

$$\delta^t_{jm} = \beta^t_p x_{jm} + \xi^t_{jm}$$

If choices by type $\mathbb{P}^t_{jm} = \mathbb{P}_m(j|\omega^t)$ observed are, then:

$$\log \mathbb{P}^t_{jm} = \beta^{k^t(j)} x_{jm} + \xi_{jm} + (\sigma_{k^t(j)} - 1) \log IV^t_{k^t(j),m} - \log IV^t_m$$

Individual heterogeneity with observed conditional shares

Assume heterogeneity can be described by types t (e.g. income quintiles) and utility given by:

$$\delta_{jm}^{t} = \beta_{p}^{t} x_{jm} + \xi_{jm}^{t}$$

If choices by type $\mathbb{P}_{jm}^{t} = \mathbb{P}_{m}(j|\omega^{t})$ observed are, then:

$$\log \mathbb{P}_{jm}^{t} = \beta^{k^{t}(j)} x_{jm} + \xi_{jm} + (\sigma_{k^{t}(j)} - 1) \log IV_{k^{t}(j),m}^{t} - \log IV_{m}^{t}$$

Two cases:

1. No common nesting structures: classify even type-by-type
2. Common nesting structures: solve joint problem

Outline

Higher-level Nesting Structures

Upstream and downstream nests given by $A_1, \ldots, A_N$ and $B_1, \ldots, B_K$, respectively.

Higher-level Nesting Structures

Upstream and downstream nests given by $A_1, ..., A_N$ and $B_1, ..., B_K$, respectively.

Log choice probabilities can be written as:

$$\log \mathbb{P}_{jm} = \beta^k x_{jm} + (\sigma^k - 1) \log IV_m^k + (\sigma^n - 1) \log IV_m^n - \log IV_m$$

so that

$$\log \mathbb{P}_{jm} = \beta^k x_{jm} + \lambda_m^k + \lambda_m^n$$

Higher-level Nesting Structures

Upstream and downstream nests given by $A_1, \ldots, A_N$ and $B_1, \ldots, B_K$, respectively.

Log choice probabilities can be written as:

$$\log \mathbb{P}_{jm} = \beta^k x_{jm} + (\sigma^k - 1) \log IV_m^k + (\sigma^n - 1) \log IV_m^n - \log IV_m$$

so that

$$\log \mathbb{P}_{jm} = \beta^k x_{jm} + \lambda_m^k + \lambda_m^n$$

We can run same classification algorithm!
→ Note: needs normalization of some $\lambda_m^k = 0$ to avoid co-linear group-market fixed effects.

Empirical application: Demand for beer

Nielsen IQ data

Data Description:

- We define products as Universal Product Codes (UPCs)
- Focus on UPCs categorized as beer (24,188 unique UPCs)
- Define markets as states (exclude Alaska and Hawaii)
- Each observation contains data on:
  - Total number of sales by UPC and state
  - (Weighted average) Prices
  - Product characteristics: unit quantity, total units, type of beer, brand, packaging, domestic

Nielsen IQ data

Data Description:

- We define products as Universal Product Codes (UPCs)
- Focus on UPCs categorized as beer (24,188 unique UPCs)
- Define markets as states (exclude Alaska and Hawaii)
- Each observation contains data on:
    - Total number of sales by UPC and state
    - (Weighted average) Prices
    - Product characteristics: unit quantity, total units, type of beer, brand, packaging, domestic

Unbalanced panel with 2,806 unique UPCs (11.6%) and 64,497 observations:

- Keep products that are sold in at least ten states
- Cover between 61.5% to 99% of volume sales across states
- Average UPC is sold across 23.32 states $\sim$ 50% missing markets

Summary statistics

Panel A: Full Sample Summary Statistics					
Variable	Mean	Std. Dev.	Median	Min	Max
Price per ml ($)	0.0063	0.0053	0.0049	0.0002	0.0519
Price ($)	10.8193	5.8026	10.7381	1.6338	25.6800
Unit quantity (units)	5.9316	5.8582	6.0000	1.0000	36.0000
Unit size (ml)	607.8046	856.3349	354.8820	207.0145	8517.1680

Panel B: Median Characteristics of the Top Ten National Brands					
Brand	Market share	Unit price/ml	Unit price	Unit quantity	Unit size (ml)
Bud Light	13.79%	0.004	9.54	6.00	354.88
Modelo	8.91%	0.004	9.90	4.00	354.88
Miller	7.92%	0.003	12.31	12.00	354.88
Coors	7.43%	0.003	12.68	12.00	354.88
Michelob	7.03%	0.004	13.26	7.00	354.88
Bud	6.71%	0.003	8.43	6.00	473.18
Corona	6.03%	0.004	11.24	6.00	354.88
Busch	3.62%	0.003	10.87	12.00	354.88
Natural	3.25%	0.002	10.74	12.00	354.88
New Belgium	2.53%	0.005	10.81	6.00	354.88

Empirical Model

We model the average utility δ_{jm} as follows:

$$\delta_{jm} = \beta_p \log \text{price}_{jm} + \beta_s \log \text{size}_j + \beta_q \log \text{quantity}_j + \xi_{jm}$$

Instrument for prices:

- Gandhi-Houde instruments:

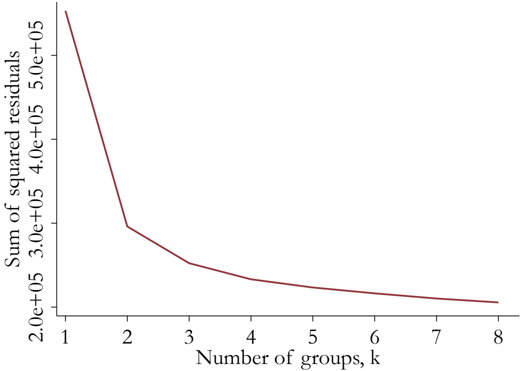
$$\text{gh}^1_{jm} = \sum_{j' \notin \text{Brand}(j,m)} (\log \text{size}_j - \log \text{size}'_j)^2 \quad \text{and} \quad \text{gh}^2_{jm} = \sum_{j' \notin \text{Brand}(j,m)} (\log \text{quantity}_j - \log \text{quantity}'_j)^2,$$

- Hausman instrument:

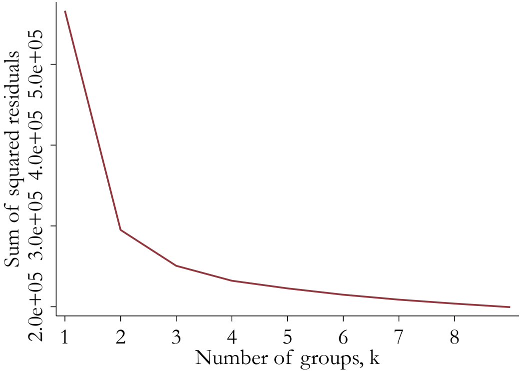
$$\text{h}_{jm} = \frac{1}{N_j - 1} \sum_{m' \neq m} \log \text{p}_{jm'},$$

First-step: Choosing the number of groups

Figure: Sum of squared residuals across different number of groups



(a) Gandhi-Houde instruments



(b) Hausman instruments

First-step: Group characteristics

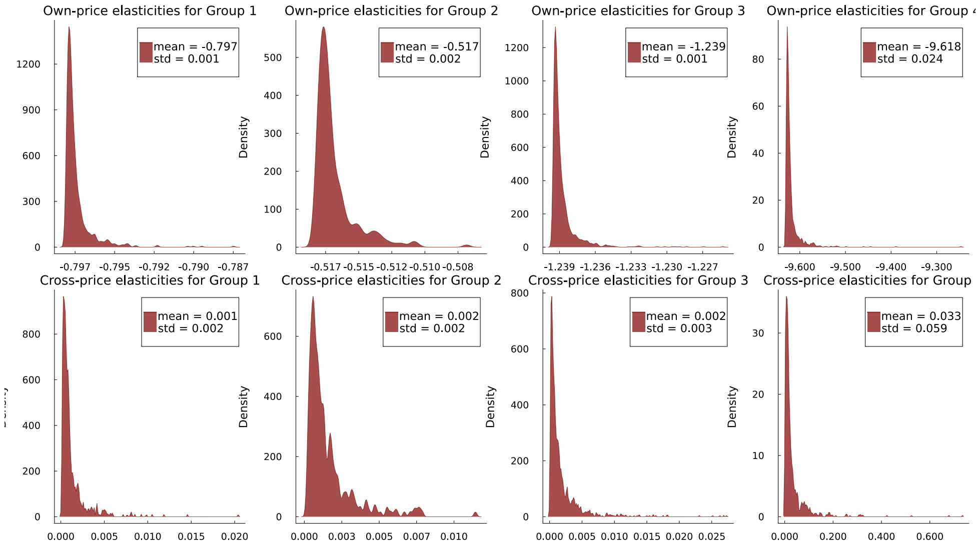
Characteristics	Unconditional median values	Conditional median values by group			
		1	2	3	4
Unit size (ml)	354.882	354.882	354.882	354.882	354.882
Unit quantity	6.000	6.000	6.000	6.000	4.000
Unit price	10.738	10.346	10.209	11.205	10.782
Price per ml	0.005	0.005	0.004	0.005	0.006
Share of domestic beer	0.751	0.793	0.802	0.737	0.719
Share of ale	0.421	0.394	0.219	0.449	0.466
Share of regular beer	0.335	0.366	0.380	0.324	0.311
Share of stout and porter	0.079	0.050	0.029	0.077	0.121
Share of light beer	0.123	0.140	0.343	0.103	0.068
Share of other types	0.042	0.050	0.029	0.047	0.034
Share of top ten brands	0.170	0.207	0.579	0.108	0.101
# of products	2806	658	242	1103	803
Label		Regular Lager	Light Beer	Regular Non-Lager	Craft Beer

Second-step results

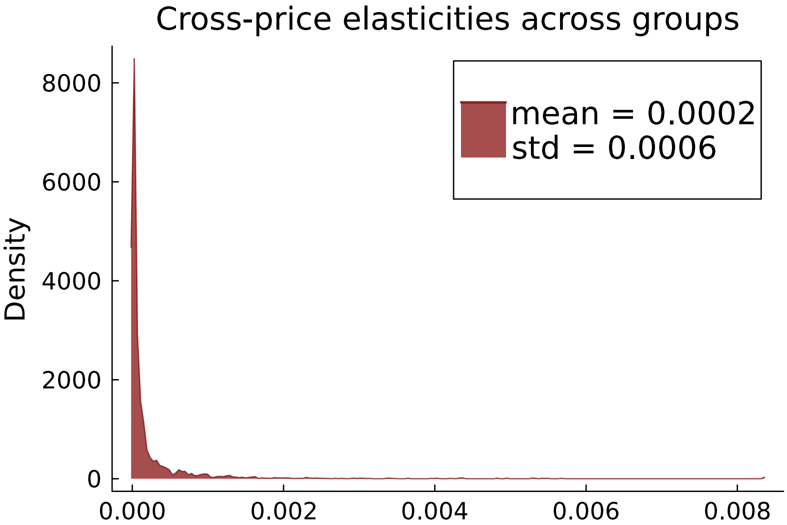
Table: Estimation Results for Nested Logit with $k^* = 4$

	OLS	IV
log unit size	0.070 (0.006)	0.340 (0.025)
log upc quantity	0.080 (0.005)	0.443 (0.027)
log price	-0.054 (0.006)	-0.419 (0.034)
σ_1	0.142 (0.001)	0.525 (0.014)
σ_2	0.343 (0.002)	0.809 (0.016)
σ_3	-0.015 (0.001)	0.338 (0.013)
σ_4	-0.320 (0.001)	0.044 (0.015)
Own-price elasticity	1.357 (1.808)	-3.471 (3.900)
Cross-price elasticity	-0.001 (0.006)	0.003 (0.019)
Market Fixed Effects	✓	✓
IV Type		gh_1, gh_2
Number of Products	2806	
Number of Observations	64497	

Substitution patterns: Within-group price elasticities



Substitution Patterns: Cross-group Cross-price Elasticities



Toward more flexible substitution patterns: Nested logit vs. mixed logit

Mixed Logit: Logit + Random coefficients in preferences with heterogeneity across consumers

+ Most flexible model in terms of substitution patterns
  $\longrightarrow$ Nested Logit is a particular case of Mixed Logit with group dummies

Toward more flexible substitution patterns: Nested logit vs. mixed logit

Mixed Logit: Logit + Random coefficients in preferences with heterogeneity across consumers

+ Most flexible model in terms of substitution patterns
  → Nested Logit is a particular case of Mixed Logit with group dummies
- Non-linear estimation: numerical integration, no closed-form demand, numerical instability
  → Dube, Fox and Su (2012), Knittel and Metaxoglou (2014)
- Cannot capture substitution on unobservable components of utility
  → e.g. neighborhood choice and preferences for gentrification

Toward more flexible substitution patterns: Nested logit vs. mixed logit

Mixed Logit: Logit + Random coefficients in preferences with heterogeneity across consumers

+ Most flexible model in terms of substitution patterns
  → Nested Logit is a particular case of Mixed Logit with group dummies
- Non-linear estimation: numerical integration, no closed-form demand, numerical instability
  → Dube, Fox and Su (2012), Knittel and Metaxoglou (2014)
- Cannot capture substitution on unobservable components of utility
  → e.g. neighborhood choice and preferences for gentrification

Data Driven Nested Logit:

- More restrictive substitution patterns

Toward more flexible substitution patterns: Nested logit vs. mixed logit

Mixed Logit: Logit + Random coefficients in preferences with heterogeneity across consumers

+ Most flexible model in terms of substitution patterns
  → Nested Logit is a particular case of Mixed Logit with group dummies
- Non-linear estimation: numerical integration, no closed-form demand, numerical instability
  → Dube, Fox and Su (2012), Knittel and Metaxoglou (2014)
- Cannot capture substitution on unobservable components of utility
  → e.g. neighborhood choice and preferences for gentrification

Data Driven Nested Logit:

- More restrictive substitution patterns
+ Closed form solutions for choice probability → Useful for a large set of applications
+ Can capture data-driven dimensions of substitution that are ex-ante unobservable through grouping structure → e.g. neighborhood choice and data-driven way of defining gentrification

Empirical comparison with mixed logit model

We estimate the following mixed-logit model using PyBLP Conlon and Gortmaker (2020):

$$u_{ijm} = \delta_{jm} + \mu_{ijm} + \epsilon_{ijm},$$

where

$$\delta_{jm} = \beta_p \text{log prize}_{jm} + \beta_s \text{log size}_j + \beta_q \text{log quantity}_j + \xi_{jm},$$

and μ_{ijm} is drawn from

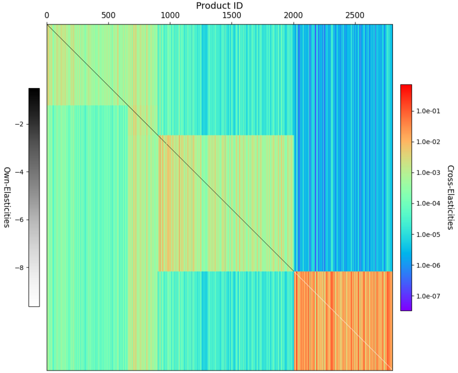
$$\mu_{ijm} = \left(\sigma_p \nu_i\right) \log \text{p}_{jm},$$

where σ_p governs the variance of the idiosyncratic component and $\nu_i \sim N(0, 1)$.

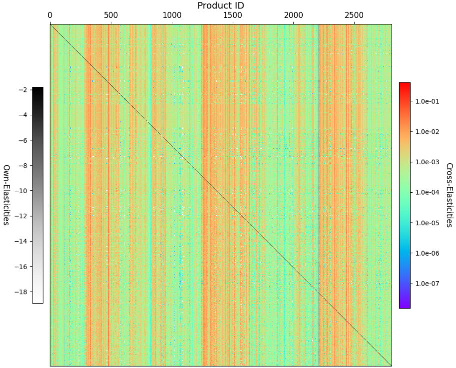
Mixed-Logit estimation results

Panel A: Linear Coefficients	β
log price	-3.775 (0.090)
log unit size	3.077 (0.062)
log unit quantity	3.222 (0.062)
Panel B: Non-linear Coefficients	σ_p
log price	0.002 (0.0003)
Mean own-price elasticity	-4.683 (2.863)
Mean cross-price elasticity	0.002 (0.005)
Market Fixed Effects	✓
IV Type	Gandhi-Houde
Number of Products	2806
Number of Observations	64497

Comparison: Matrix of price elasticities



(a) Nested logit ($k^* = 4$)



(b) Mixed Logit

Conclusions and next steps

Conclusions:

- Proposed a two-step estimator to estimate nesting structure
- Showed conditions for consistency of estimator
- Empirical application based on US beer market shows reasonable results and rich substitution patterns

Conclusions and next steps

Conclusions:

- Proposed a two-step estimator to estimate nesting structure
- Showed conditions for consistency of estimator
- Empirical application based on US beer market shows reasonable results and rich substitution patterns

Next steps:

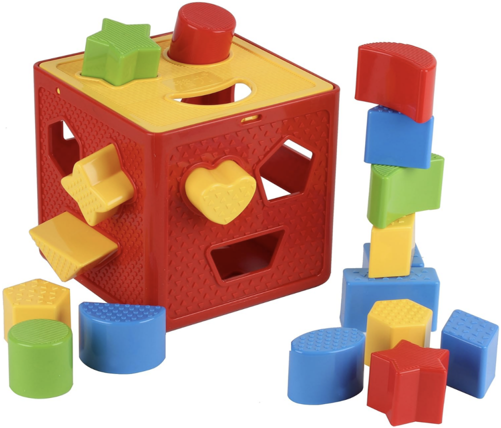
- Show consistency in other nested demand structures (e.g. Type II EV, nested CES demand)
- Extend empirical applications
  - Demand for beer: Add consumer heterogeneity by linking to data on demographics (ACS)
  - Spatial:
    - Labor markets clusters: Is NYC a closer substitute to SF or Newark?
    - Defining market structure for spatial applications: what's a neighborhood?
  - Trade: Revisit CES nested demand a lá Broda and Weisnstein (2006)

Thanks!

Appendix

Why asymptotics on M?

Graphical toy example: playing in the dark



- You can't see
- Different shapes are placed in different urns
- Every minute m, you can learn by touching an unknown shape from an urn for 5 seconds
- Your goal is to get as many shapes through the box as possible
- The more minutes m, the more you learn!

Why asymptotics on M?

Simplified example

Back

- Consider the following simplified model with $G = 2$:

$$y_{im} = \alpha^*_{k_i^*} + v_{im}, \qquad k_i \in \{1, 2\}.$$

- We characterize the misclassification probability:

$$\Pr\left(\widehat{k}_i(\alpha) = 2 \middle| k_i^* = 1\right) = \Pr\left((\bar{y}_i - \alpha_2)^2 < (\bar{y}_i - \alpha_1)^2 \middle| k_i^* = 1\right).$$

- If v_{im} are iid normal $(0, \sigma^2)$ and $\alpha_1 < \alpha_2$ then this is:

$$\Pr\left(\overline{v}_i > \frac{\alpha_1 + \alpha_2}{2} - \alpha_1^*\right) \quad = \quad 1 - \Phi\left(\frac{\sqrt{M}}{\sigma}\left(\frac{\alpha_1 + \alpha_2}{2} - \alpha_1^*\right)\right),$$

  which vanishes exponentially fast as M increases.

- **Intuition:** When M grows
  - Mean $\bar{y}_i$ converges to α_1
  - If mis-classified, then error $\bar{y}_i - \alpha_2$ eventually should become larger than $\bar{y}_i - \alpha_1$
  - Every m is a chanve to learn: With enough opportunities, we'll eventually learn the truth!