Data-Driven Nests

FGV

December 4th, 2025

Milena Almagro
(Chicago Booth & NBER)

Kenneth Lai
(LMU Munich)

Elena Manresa
(Princeton University)

Demand Estimation

Discrete Choice Models

- Models of discrete choice are the workhorse in demand estimation

Demand Estimation

Discrete Choice Models

- Models of discrete choice are the workhorse in demand estimation
- Random utility framework:

Demand Estimation

Discrete Choice Models

- Models of discrete choice are the workhorse in demand estimation
- Random utility framework:
  - Utility is driven by observables + idiosyncratic taste shock
  - Agents choose the alternative with highest utility

Demand Estimation

Discrete Choice Models

- Models of discrete choice are the workhorse in demand estimation

- Random utility framework:
  - Utility is driven by observables + idiosyncratic taste shock
  - Agents choose the alternative with highest utility

- When shocks $\sim$ Type I EV $\implies$ Multinomial logit

Demand Estimation

Discrete Choice Models

- Models of discrete choice are the workhorse in demand estimation
- Random utility framework:
  - Utility is driven by observables + idiosyncratic taste shock
  - Agents choose the alternative with highest utility
- When shocks $\sim$ Type I EV $\implies$ Multinomial logit
  - Closed form solutions of choice probability + low number of parameters
  - Generates unrealistic substitution patterns (IIA) Blue-bus/Red-bus

Demand Estimation

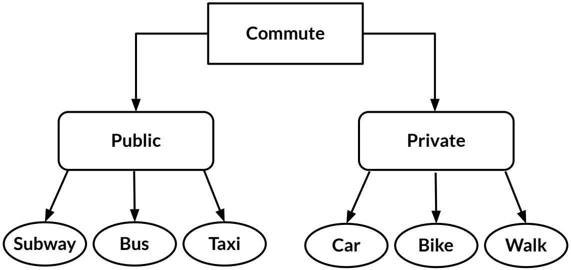
Discrete Choice Models

- Models of discrete choice are the workhorse in demand estimation
- Random utility framework:
  - Utility is driven by observables + idiosyncratic taste shock
  - Agents choose the alternative with highest utility
- When shocks $\sim$ Type I EV $\implies$ Multinomial logit
  - Closed form solutions of choice probability + low number of parameters
  - Generates unrealistic substitution patterns (IIA) Blue-bus/Red-bus
- A common alternative is Nested Logit (NL)
  - Natural extension of Multinomial Logit

Main features of Nested Logit models

- Nests lead to different degrees of substitution:
  $\longrightarrow$ More substitution **within** than **across** nests
- Within-nest correlation of idiosyncratic shocks
- Interpretable model as "sequential" choice
- Low number of parameters
- Closed form solutions of choice probabilities
  $\longrightarrow$ Easy to implement

McFadden (1978)



Nested demand models are ubiquitous across many fields

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Ferreira, Petrin & Waldfogel (2026), Björnerstedt & Verboven (2016), Atalay et al. (2025)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Ferreira, Petrin & Waldfogel (2026), Björnerstedt & Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Khandelwal (2010), Costinot, Donaldson & Komunjer (2012), Caliendo & Parro (2015), Allcott et al. (2024)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Ferreira, Petrin & Waldfogel (2026), Björnerstedt & Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Khandelwal (2010), Costinot, Donaldson & 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)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Ferreira, Petrin & Waldfogel (2026), Björnerstedt & Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Khandelwal (2010), Costinot, Donaldson & 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)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Ferreira, Petrin & Waldfogel (2026), Björnerstedt & Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Khandelwal (2010), Costinot, Donaldson & 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 & Burstein (2008), Edmond, Midrigan & Xu (2015), Berger, Herkenhoff & Mongey (2022)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Ferreira, Petrin & Waldfogel (2026), Björnerstedt & Verboven (2016), Atalay et al. (2025)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Khandelwal (2010), Costinot, Donaldson & 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 & Burstein (2008), Edmond, Midrigan & Xu (2015), Berger, Herkenhoff & Mongey (2022)...

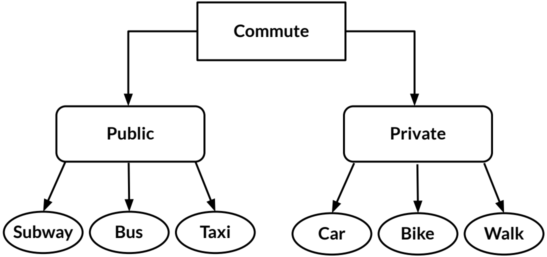
$\Longrightarrow$ In many applications, demand can be appropriately described by a nesting structure

What's the right nesting structure?

In general, the practitioner faces the challenge of specifying the nesting structure *ex-ante*.

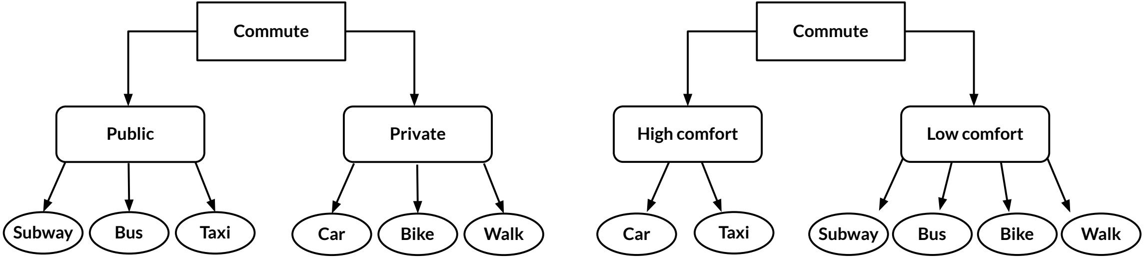
What's the right nesting structure?

In general, the practitioner faces the challenge of specifying the nesting structure *ex-ante*.



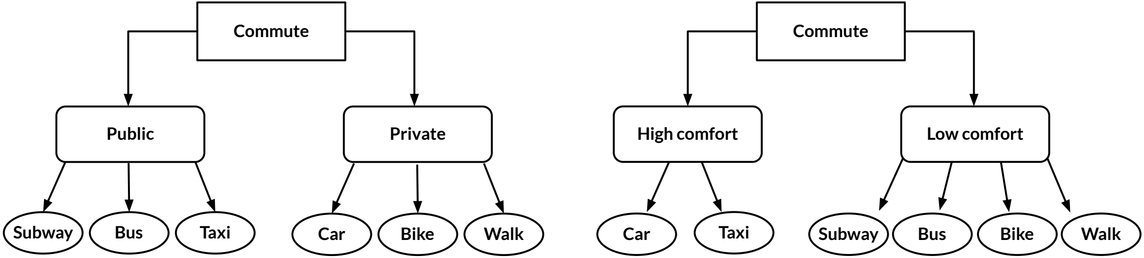
What's the right nesting structure?

In general, the practitioner faces the challenge of specifying the nesting structure *ex-ante*.



What's the right nesting structure?

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. Estimate nests with k-means clustering on demand curves (Bonhomme and Manresa, 2015):
     - Nested Logit is a correlated shock that leads to group-specific substitution patterns and elasticities
     - Demand curves with group-specific coefficients **common within nests** but that **vary across nests**
  2. Estimate model parameters as if the groups were known (Berry, 1994)
- We exploit the structure of the model, the availability of many markets and of many products
- Empirical application: Demand for beer
  - In this setting, data-driven model outperforms Naive Nested Logit and Mixed Logit out of sample

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

1. Empirical model
   1.1 Identification
   1.2 Estimation
2. Practical aspects and extensions
   2.1 Choosing the number of groups
   2.2 Endogenous prices
   2.3 Introducing consumer heterogeneity
   2.4 Higher-level nesting structures
3. Consistency, statistical properties, and Monte Carlo
4. Empirical application: Demand for beer
   4.1 Data description and empirical model
   4.2 First-step estimation: Classification
   4.3 Second-step estimation: Structural parameters
   4.4 Comparison with alternative empirical models

Empirical model

Logit

- Consider the random 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 utility:

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

Logit

- Consider the random 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 utility:

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

- If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \exp\Big(-\sum_j e^{-\varepsilon_{ijm}}\Big)$, then:

$$\mathbb{P}_{jm} = \frac{e^{\delta_{jm}}}{\sum_{j'=1}^{J} e^{\delta_{j'm}}}$$

Nested Logit

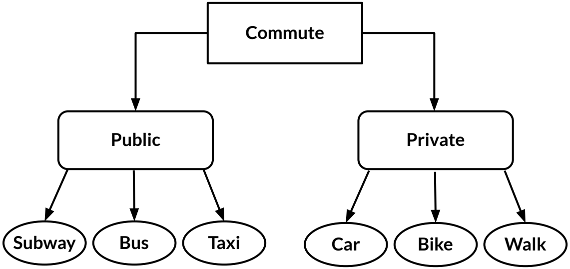
If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \exp\Big(-\sum_{k=1}^{K}\Big(\sum_{j\in k} e^{-\varepsilon_{ijm}/\sigma_k}\Big)^{\sigma_k}\Big)$, then

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

Nested Logit

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \exp\Big(-\sum_{k=1}^{K}\Big(\sum_{j\in k} e^{-\varepsilon_{ijm}/\sigma_k}\Big)^{\sigma_k}\Big)$, then

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

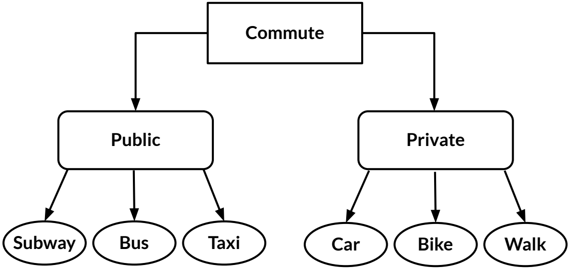


Nested Logit

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \exp\Big(-\sum_{k=1}^{K}\big(\sum_{j\in k} e^{-\varepsilon_{ijm}/\sigma_k}\big)^{\sigma_k}\Big)$, then

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

– Nest structure ~ common within-nest shock
  $\rightarrow \varepsilon_{ijm} = \eta_{k(j),m} + \tilde{\varepsilon}_{ijm}$ (Cardell, 1997)

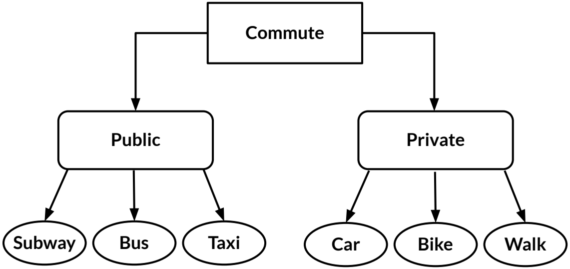


Nested Logit

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \exp\Big(-\sum_{k=1}^{K}\big(\sum_{j\in k} e^{-\varepsilon_{ijm}/\sigma_k}\big)^{\sigma_k}\Big)$, then

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

- Nest structure $\sim$ common within-nest shock
  $\rightarrow \varepsilon_{ijm} = \eta_{k(j),m} + \tilde{\varepsilon}_{ijm}$ (Cardell, 1997)
- Nests lead to different degrees of substitution:
  $\rightarrow$ More substitution **within** than **across** nests

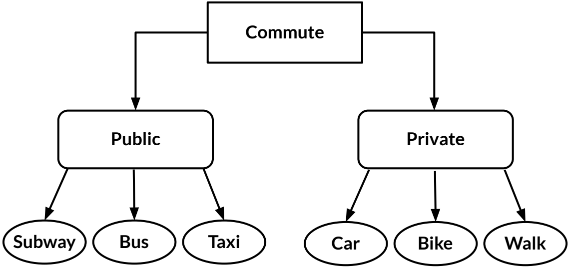


Nested Logit

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim} \exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in k} e^{-\varepsilon_{ijm}/\sigma_k}\right)^{\sigma_k}\right)$, then

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

- Nest structure ∼ common within-nest shock
  $\rightarrow \varepsilon_{ijm} = \eta_{k(j),m} + \tilde{\varepsilon}_{ijm}$ (Cardell, 1997)
- Nests lead to different degrees of substitution:
  → More substitution **within** than **across** nests
- $\sigma^{k(j)} \in (0, 1]$ captures within-nest independence
  → if $\sigma^{k(j)} = 1 \implies$ back to logit



Towards an empirical equation

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

Towards an empirical equation

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

Towards an empirical equation

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} = e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \Lambda_m^{k(j)} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma^{k(j)}} + \lambda_m^{k(j)}$$

Towards an empirical equation

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} = e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \Lambda_m^{k(j)} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma^{k(j)}} + \lambda_m^{k(j)}$$

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

Towards an empirical equation

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} = e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \Lambda_m^{k(j)} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma^{k(j)}} + \lambda_m^{k(j)}$$

Assume $\delta_{jm} = \beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}$. With a slight abuse of notation for ξ_{jm}, it follows:

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

Towards an empirical equation

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma^l}} \right)^{\sigma^l}} = e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \Lambda_m^{k(j)} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma^{k(j)}} + \lambda_m^{k(j)}$$

Assume $\delta_{jm} = \beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}$. With a slight abuse of notation for ξ_{jm}, it follows:

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

Key observations: $(\beta_x^{k(j)}, \beta_p^{k(j)}, \lambda_m^{k(j)})$ common for products in the same group!

- Common $(\beta_x^{k(j)}, \beta_p^{k(j)})$ implies marginal effects are **common within nests** and **differ across nests**
  $\rightarrow$ Captures different elastiticites of substitution across nests
- Conditional on covariates, $\lambda_m^{k(j)}$ is a **correlated shock within nest** that can vary across markets
  $\rightarrow$ Example: craft beer has become more popular in the Midwest and coasts relative to the South

Outline

Intuition on nest identification

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

Intuition on nest identification

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

For simplicity, let's consider a model where $\mathbb{E}[\xi_{jm} | x_{jm}, p_{jm}] = 0$, and take means across markets:

$$\overline{\log \mathbb{P}_j} = \beta_x^{k(j)} \bar{x}_j + \beta_p^{k(j)} \overline{p_j} + \overline{\lambda}^{k(j)}, \text{ where } \overline{\xi_j} \to 0$$

Intuition on nest identification

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

For simplicity, let's consider a model where $\mathbb{E}[\xi_{jm} | x_{jm}, p_{jm}] = 0$, and take means across markets:

$$\overline{\log \mathbb{P}_j} = \beta_x^{k(j)} \bar{x}_j + \beta_p^{k(j)} \overline{p_j} + \overline{\lambda}^{k(j)}, \text{ where } \overline{\xi_j} \to 0$$

Key observation: There are K lines with coefficients $(\overline{\lambda}^{k(j)}, \beta_x^{k(j)}, \beta_p^{k(j)})$ that perfectly fit the data

Intuition on nest identification

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

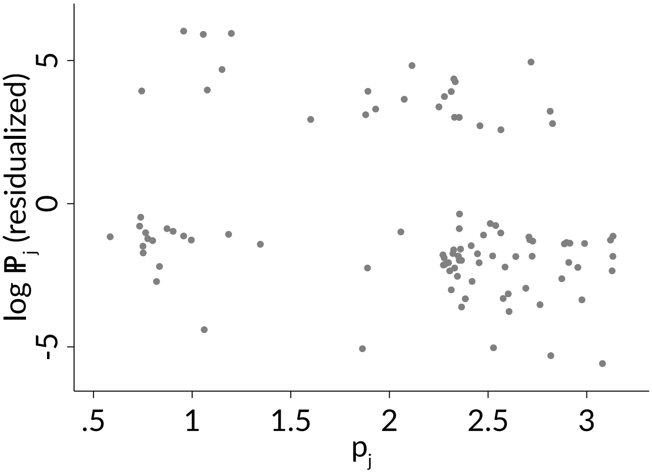
For simplicity, let's consider a model where $\mathbb{E}[\xi_{jm}|x_{jm}, p_{jm}] = 0$, and take means across markets:

$$\overline{\log \mathbb{P}_j} = \beta_x^{k(j)} \bar{x}_j + \beta_p^{k(j)} \overline{p_j} + \overline{\lambda}^{k(j)}, \text{ where } \overline{\xi_j} \to 0$$

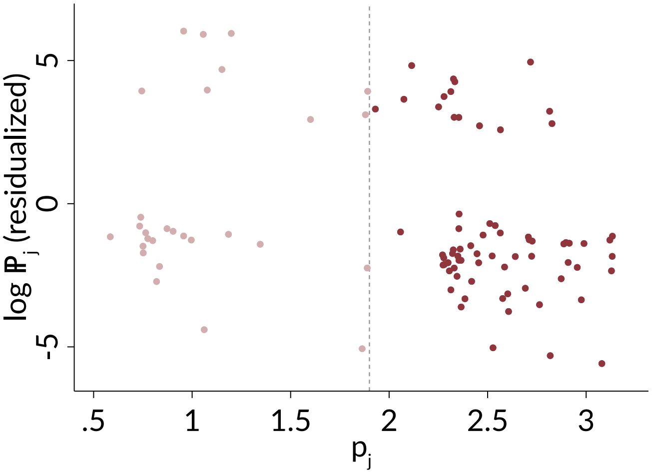
Key observation: There are K lines with coefficients $(\overline{\lambda}^{k(j)}, \beta_x^{k(j)}, \beta_p^{k(j)})$ that perfectly fit the data

$\to$ Not only search on coefficients, $(\beta^1, ..., \beta^K, \lambda^1, ..., \lambda^K)$ but also on how to partition data

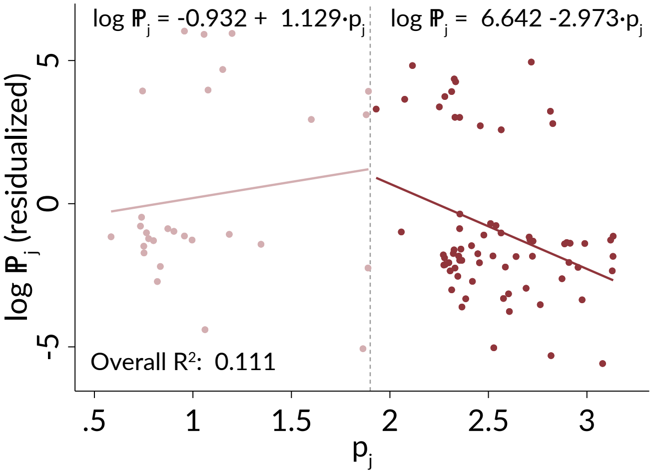
Intuition on nest identification: graphic example with $K = 2$



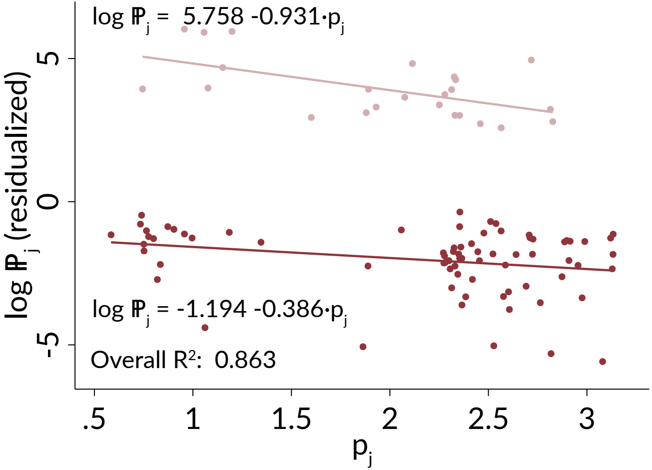
Intuition on nest identification: graphic example with $K = 2$



Intuition on nest identification: graphic example with $K = 2$



Intuition on nest identification: graphic example with $K = 2$



Outline

Estimation of groups

First-step: Classification

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

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

→ Equivalent to k-means but using regressions instead of cluster means

Estimation of groups

First-step: Classification

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

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

$\rightarrow$ Equivalent to k-means but using regressions instead of cluster means

Combinatorial, non-convex problem!

$\rightarrow$ Solution: iterative clustering algorithm extending Bonhomme and Manresa (2015)

Two-step strategy

First Step: Classification

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

Two-step strategy

First Step: Classification

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

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

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

Two-step strategy

First Step: Classification

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

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

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

3. For new grouping structure $\mathcal{B}^{s+1}$, compute:

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

Two-step strategy

First Step: Classification

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

$$k(j)^{s+1} = \underset{k \in \{1,\ldots,K\}}{\arg\min} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_p^{k,s} p_{jm} + \beta_x^{k,s} x_{jm} + \lambda_m^{k,s}) \right)^2,$$

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

3. For new grouping structure $\mathcal{B}^{s+1}$, compute:

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

4. Repeat until convergence of parameters.
5. In practice, to find global optimum, repeat for different starting values and take minimum

Two-step strategy

Second step: Estimation of structural parameters

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

Two-step strategy

Second step: Estimation of structural parameters

Once we have grouping structure $\{k(1), ..., k(J)\}$, estimate β and σ^k 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 of structural parameters

Once we have grouping structure $\{k(1), \ldots, k(J)\}$, estimate β and σ^k 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 sets of instruments:

- Instrument for prices, z_{jm}
- Instrument for $\log \mathbb{P}_{j,m|k(j)} = \frac{e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma^{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma^{k(j)}}}} \implies$ Functions of $x_{j'm}$ for $j' \in k(j)$

Practical aspects and 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 $k \in \{1, ..., K\}$:

- Divide products into N equal subsets, $P_1, ..., P_N$.
- 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

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

Outline

Endogenous prices

Exogeneity of characteristics would imply:

$$\mathbb{E}[\xi_{jm} | x_{jm}, p_{jm}] = 0$$

$\rightarrow$ could be unrealistic for many applications, especially for prices

Endogenous prices

Exogeneity of characteristics would imply:

$$\mathbb{E}[\xi_{jm} | x_{jm}, p_{jm}] = 0$$

$\rightarrow$ could be unrealistic for many applications, especially for prices

In the presence of exogenous shifter z_{jm}, consider for now a linear price projection:

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

Endogenous prices

Exogeneity of characteristics would imply:

$$\mathbb{E}[\xi_{jm}|x_{jm}, p_{jm}] = 0$$

$\rightarrow$ could be unrealistic for many applications, especially for prices

In the presence of exogenous shifter z_{jm}, consider for now a linear price projection:

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

Substitute into main equation:

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

More general pricing equations

Throughout the talk we will take:

$$p_{jm} = \gamma_x x_{jm} + \gamma_z z_{jm} + v_{jm}$$

More general pricing equations

Throughout the talk we will take:

$$p_{jm} = \gamma_x x_{jm} + \gamma_z z_{jm} + v_{jm}$$

But method accommodates more general pricing models:

- Non-linear functions:

$$p_{jm} = f(x_{jm}, z_{jm}, v_{jm})$$

More general pricing equations

Throughout the talk we will take:

$$p_{jm} = \gamma_x x_{jm} + \gamma_z z_{jm} + v_{jm}$$

But method accommodates more general pricing models:

- Non-linear functions:

$$p_{jm} = f(x_{jm}, z_{jm}, v_{jm})$$

- Group-specific coefficients:

$$p_{jm} = \gamma_x^{g(j)} x_{jm} + \gamma_z^{g(j)} z_{jm} + v_{jm}$$

More general pricing equations

Throughout the talk we will take:

$$p_{jm} = \gamma_x x_{jm} + \gamma_z z_{jm} + v_{jm}$$

But method accommodates more general pricing models:

- Non-linear functions:

$$p_{jm} = f(x_{jm}, z_{jm}, v_{jm})$$

- Group-specific coefficients:

$$p_{jm} = \gamma_x^{g(j)} x_{jm} + \gamma_z^{g(j)} z_{jm} + v_{jm}$$

- Non-linear functions of x_{jm} and z_{jm} with groupings:

$$p_{jm} = f^{g(j)}(x_{jm}, z_{jm}) + v_{jm}$$

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 $A_1, \ldots, A_N$ 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

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

Higher-level Nesting Structures

Upstream and downstream nests given by $A_1, \dots, A_N$ and $B_1, \dots, B_K$, 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!
$\rightarrow$ Note: needs normalization of some $\lambda_m^k = 0$ to avoid co-linear group-market fixed effects.

Consistency, statistical properties, and Monte Carlo

Regularity conditions with unknown nests

1. K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$
   - Need to observe market shares multiple times so error ξ_{jm} vanishes out Why $M \to \infty$?

Regularity conditions with unknown nests

1. K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$
   - Need to observe market shares multiple times so error ξ_{jm} vanishes out Why $M \to \infty$?
2. Orthogonality of errors: $\mathbb{E}[\xi_{jm}, v_{jm} | x_{jm}, z_{jm}] = 0$
3. Variation in x_{jm}, z_{jm} conditional on groups and thin tails

Regularity conditions with unknown nests

1. K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$
   - Need to observe market shares multiple times so error ξ_{jm} vanishes out Why $M \to \infty$?
2. Orthogonality of errors: $\mathbb{E}[\xi_{jm}, v_{jm} | x_{jm}, z_{jm}] = 0$
3. Variation in x_{jm}, z_{jm} conditional on groups and thin tails
4. $\lambda_m^{k(j)}$ is a random variable: Conditions so that they are well behaved when $J \to \infty$

Regularity conditions with unknown nests

1. K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$
   - Need to observe market shares multiple times so error ξ_{jm} vanishes out Why $M \to \infty$?
2. Orthogonality of errors: $\mathbb{E}[\xi_{jm}, v_{jm} | x_{jm}, z_{jm}] = 0$
3. Variation in x_{jm}, z_{jm} conditional on groups and thin tails
4. $\lambda_m^{k(j)}$ is a random variable: Conditions so that they are well behaved when $J \to \infty$
5. Group separation: cannot have $\beta^k = \beta^{k'}$ and $\lambda_m^k = \lambda_m^{k'}$ for all m for some $k \neq k'$
   - This requires that two groups contain same products and look *identical* across all markets

$$x_{jm} = x_j,\ z_{jm} = z_j\ \forall m$$

   - If variation in x_{jm}, z_{jm}, this event happens with probability zero.

Regularity conditions with unknown nests

1. K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$
   - Need to observe market shares multiple times so error ξ_{jm} vanishes out Why $M \to \infty$?
2. Orthogonality of errors: $\mathbb{E}[\xi_{jm}, v_{jm} | x_{jm}, z_{jm}] = 0$
3. Variation in x_{jm}, z_{jm} conditional on groups and thin tails
4. $\lambda_m^{k(j)}$ is a random variable: Conditions so that they are well behaved when $J \to \infty$
5. Group separation: cannot have $\beta^k = \beta^{k'}$ and $\lambda_m^k = \lambda_m^{k'}$ for all m for some $k \neq k'$
   - This requires that two groups contain same products and look *identical* across all markets

$$x_{jm} = x_j, \; z_{jm} = z_j \; \forall m$$

   - If variation in x_{jm}, z_{jm}, this event happens with probability zero.

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

Details

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:

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

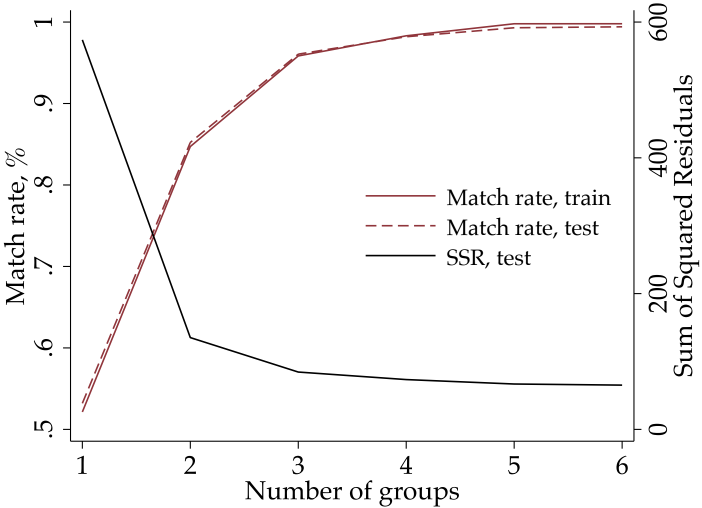
Generate data as follows:

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

Set $\rho = 0.3$

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 $M = 10$.



Results with $K = 3$

50 Bootstrap iterations

J	M	Runtime	Matched	True	β_p	β_c	σ_1	σ_2	σ_3
				True	-1	1	0.2	0.3	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	2:11:14	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

Empirical application: Demand for beer

Outline

Nielsen IQ store panel data

Data Description:

- We define products as Universal Product Codes (UPCs)
- Focus on year 2022 and on UPCs categorized as beer (24,188 unique UPCs)
- Define markets as states (exclude Alaska and Hawaii) $\implies M = 48$
- Each observation contains data on:
  - Total number of sales by UPC and state, aggregating across stores
  - Product characteristics: Prices, unit quantity, total units, type of beer, brand, packaging, domestic

Nielsen IQ store panel data

Data Description:

- We define products as Universal Product Codes (UPCs)
- Focus on year 2022 and on UPCs categorized as beer (24,188 unique UPCs)
- Define markets as states (exclude Alaska and Hawaii) $\implies M = 48$
- Each observation contains data on:
  - Total number of sales by UPC and state, aggregating across stores
  - Product characteristics: Prices, unit quantity, total units, type of beer, brand, packaging, domestic

Keep products that are sold in at least ten states

Nielsen IQ store panel data

Data Description:

- We define products as Universal Product Codes (UPCs)
- Focus on year 2022 and on UPCs categorized as beer (24,188 unique UPCs)
- Define markets as states (exclude Alaska and Hawaii) $\implies M = 48$
- Each observation contains data on:
  - Total number of sales by UPC and state, aggregating across stores
  - Product characteristics: Prices, unit quantity, total units, type of beer, brand, packaging, domestic

Keep products that are sold in at least ten states

- Unbalanced panel with $J = 2,806$ unique UPCs (11.6%) and 64,497 observations
- 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

Large variation in shares and covariates

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)	607.80	856.33	354.88	207.01	8517.17

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:

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

Empirical Model

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

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

Empirical model excludes categorical variables
$\implies$ worst-case scenario in providing ex-ante information of grouping structure

Empirical Model

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

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

Empirical model excludes categorical variables
$\implies$ worst-case scenario in providing ex-ante information of grouping structure

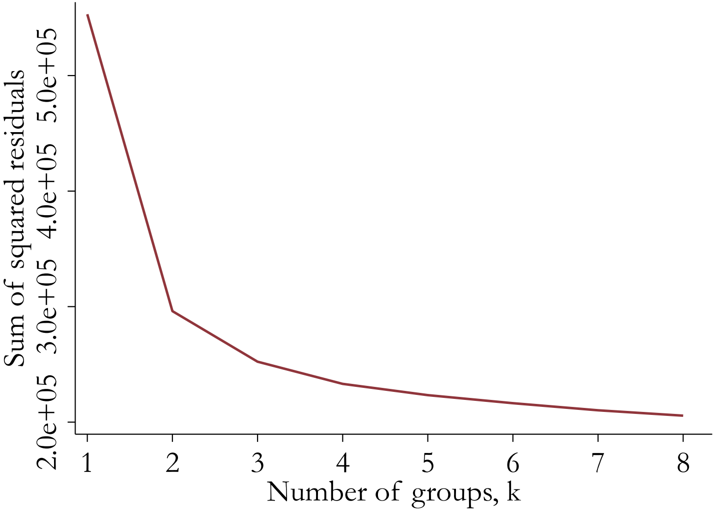
Instrument for prices: Gandhi-Houde instruments

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

Outline

First-step: Choosing the number of groups

Elbow method suggest $K^* = 4$



First-step: Group characteristics

Characteristics	Unconditional values	Conditional 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	Special /Craft	Boutique Craft

First-step: Group characteristics

Characteristics	Unconditional values	Conditional 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	Special /Craft	Boutique Craft

First-step: Group characteristics

Characteristics	Unconditional values	Conditional 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	Special /Craft	Boutique Craft

First-step: Group characteristics

Characteristics	Unconditional values	Conditional 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	Special /Craft	Boutique Craft

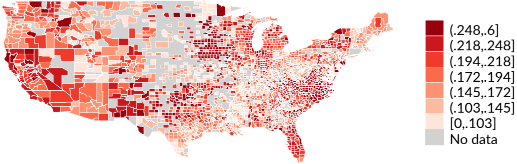
First-step: Group characteristics

Characteristics	Unconditional values	Conditional 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	Special /Craft	Boutique Craft

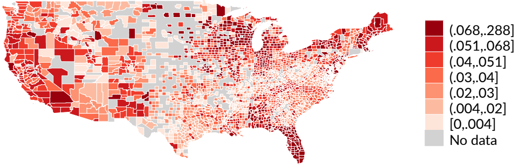
Brands by group

Spatial correlation of county market shares by group

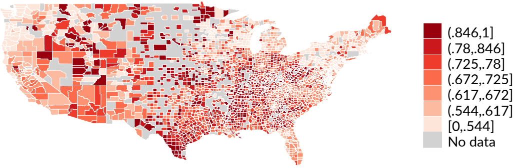
Regular lager



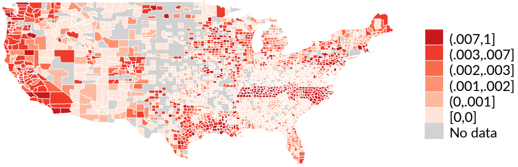
Special/craft



Light beer

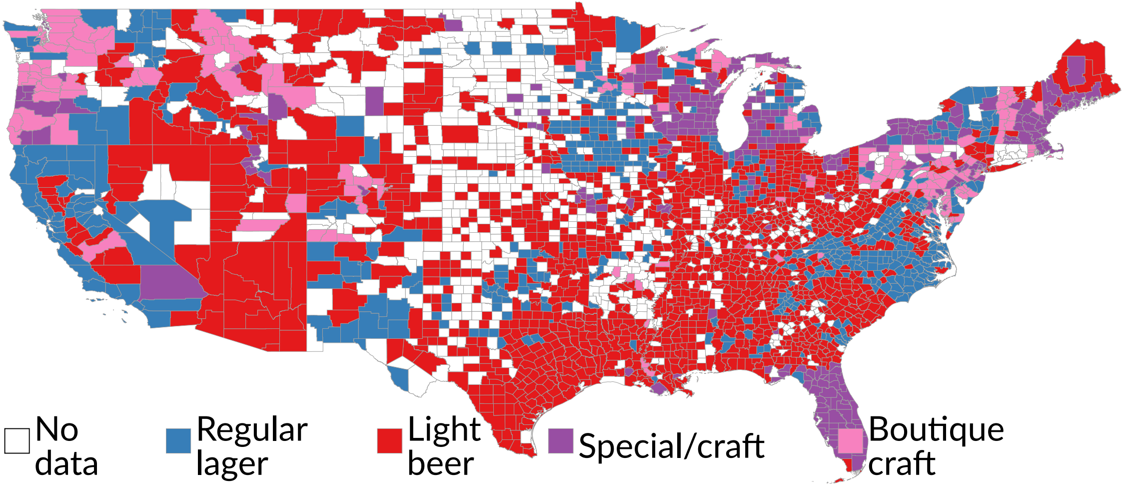


Boutique craft

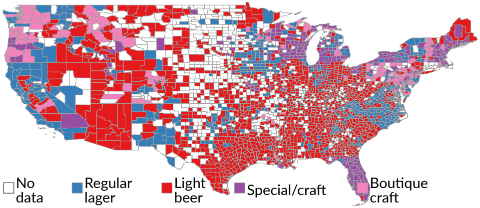


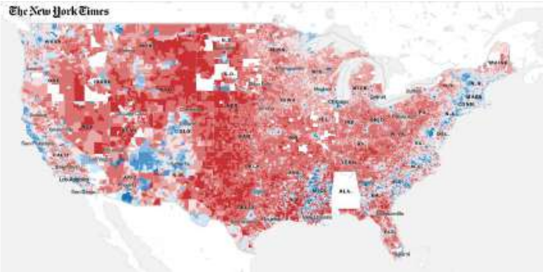
Most popular group (relative to national trends)

High degree of spatial correlation



Group popularity correlated with political outcomes!





Correlation between Light Beer share and 2024 Republican share is 0.39

Groups are capturing political preferences

This is not a mere coincidence: there are empirical studies on this!

Groups are capturing political preferences

This is not a mere coincidence: there are empirical studies on this!

Key takeaway: The Data-driven NL is also able to capture unobserved drivers of demand
(so long as they can be described or approximated by groups)

Group validation: Bud Light boycott

Testing nest substitution

In April 2023, Dylan Mulvaney—an actress, social media influencer, and transgender woman—releases a social media promotion of Bud Light

Group validation: Bud Light boycott

Testing nest substitution

The promotion sparked a backlash and calls for a boycott

Group validation: Bud Light boycott

Testing nest substitution

The boycott appeared to have some bite, and its effects were widely covered in the media

Dylan Mulvaney: Bud Light loses top spot in US after boycott

14 June 2023

Mariko Oi Business reporter

Group validation: Bud Light boycott

Testing nest substitution

Use Bud Light boycott as an exogenous negative shock to demand for Bud Light

Group validation: Bud Light boycott

Testing nest substitution

Use Bud Light boycott as an exogenous negative shock to demand for Bud Light

Hypothesis: if consumers are substituting consumption to other beers
$\implies$ Demand for beers within the same nest as Bud Light ("Light Beer") should increase more relative to beers in other nests

Group validation: Bud Light boycott

Testing nest substitution

Use Bud Light boycott as an exogenous negative shock to demand for Bud Light

Hypothesis: if consumers are substituting consumption to other beers
$\implies$ Demand for beers within the same nest as Bud Light ("Light Beer") should increase more relative to beers in other nests

Augment our state-product panel of sales by adding data for 2023

Group validation: Bud Light boycott

Testing nest substitution

Use Bud Light boycott as an exogenous negative shock to demand for Bud Light

Hypothesis: if consumers are substituting consumption to other beers
$\implies$ Demand for beers within the same nest as Bud Light ("Light Beer") should increase more relative to beers in other nests

Augment our state-product panel of sales by adding data for 2023

Empirical model

$$\log \mathbb{P}_{jmt} = \kappa_j + \eta_t + \sum_{k=1}^{K} \alpha_k \cdot \mathbb{1}\{\widehat{k}(j) = k\}\mathbb{1}\{t = 2023\} + \epsilon_{jmt}$$

Group validation: Bud Light boycott

Testing nest substitution

Use Bud Light boycott as an exogenous negative shock to demand for Bud Light

Hypothesis: if consumers are substituting consumption to other beers
$\implies$ Demand for beers within the same nest as Bud Light ("Light Beer") should increase more relative to beers in other nests

Augment our state-product panel of sales by adding data for 2023

Empirical model

$$\log \mathbb{P}_{jmt} = \kappa_j + \eta_t + \sum_{k=1}^{K} \alpha_k \cdot \mathbb{1}\{\widehat{k}(j) = k\}\mathbb{1}\{t = 2023\} + \epsilon_{jmt}$$

Because of the inclusion of UPC fixed effects κ_j

$\implies$ exploit variation (temporal changes) not used in the classification of groups

Group validation: DiD results

$$\log \mathbb{P}_{jmt} = \kappa_j + \eta_t + \sum_{k=1}^{K} \alpha_k \mathbb{1}\{\widehat{k}(j) = k\}\mathbb{1}\{t = 2023\} + \epsilon_{jmt}$$

	First stage		DiD				Placebo	
t=2023	-0.287***	(0.009)	-0.314***	(0.009)	-0.353***	(0.014)	-0.303***	(0.016)
Treatment (Budlight)	-0.156***	(0.054)						
Treatment same group			0.289***	(0.031)				
Regular beer × t=2023					0.087***	(0.025)		
Light beer × t=2023					0.328***	(0.033)		
Non-lager/craft × t=2023					0.025	(0.024)		
Boutique craft × t=2023					0.132***	(0.032)		
Placebo Regular beer × t=2023							0.020	(0.021)
Placebo Light beer × t=2023							0.020	(0.021)
Placebo Non-lager/craft × t=2023							0.030	(0.021)
Placebo Boutique craft × t=2023							0.007	(0.021)
Observations	219728		213959		208829		208829	
UPC FE	YES		YES		YES		YES	
Exclude Bud Light	NO		YES		YES		YES	
Baseline group	No Bud Light		Other groups		Outside option		Outside option	

Notes: *** p<0.01, ** p<0.05, * p<0.1

Outline

Second-step results: Data-driven Nested Logit structural parameters

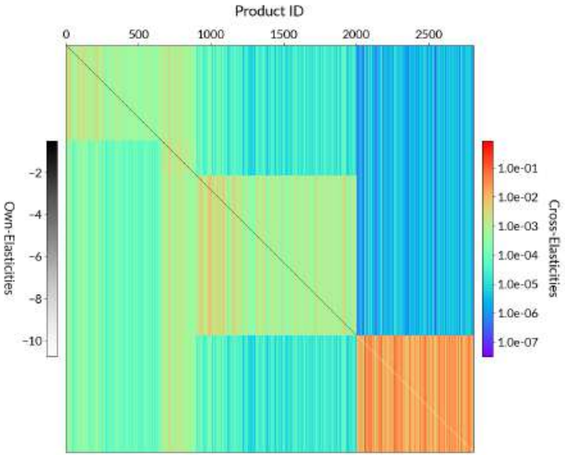
	OLS	IV
log unit size	0.070 (0.006)***	0.331 (0.025)***
log upc quantity	0.080 (0.005)***	0.432 (0.026)***
log price	-0.054 (0.006)***	-0.406 (0.033)***
$\sigma_{\text{regular/lager}}$	0.142 (0.001)***	0.520 (0.014)***
$\sigma_{\text{light beer}}$	0.343 (0.002)***	0.802 (0.016)***
σ_{craft}	-0.015 (0.001)***	0.333 (0.013)***
$\sigma_{\text{boutique craft}}$	-0.320 (0.001)***	0.038 (0.015)***
Mean Own-price Elasticity	1.357 (1.808)	-3.779 (4.414)
Mean Cross-price Elasticity	-0.001 (0.006)	0.004 (0.022)
Market Fixed Effects	✓	✓
IV Type		gh_1, gh_2
Number of Products	2806	
Number of Observations	64497	

Literature range of own-price elasticities is $[-6.3, -2.4]$

(Pinske and Slade, 2004; Slade, 2004; Romeo, 2016; Miller and Weinberg, 2017)

Data-driven Nested Logit matrix of elasticities

Rich heterogeneity in substitution patterns



Outline

Naive Nested Logit: Using NielsenIQ “Beer Type” as Nests

	OLS	IV
log unit size	-0.085 (0.008)***	1.468 (0.066)***
log upc quantity	0.090 (0.006)***	1.433 (0.066)***
log price	-0.092 (0.007)***	-1.525 (0.082)***
$\sigma_{\text{regular/lager}}$	0.110 (0.001)***	0.611 (0.015)***
$\sigma_{\text{light beer}}$	0.121 (0.001)***	0.653 (0.015)***
σ_{ale}	0.006 (0.001)***	0.568 (0.017)***
$\sigma_{\text{stout/porter}}$	-0.255 (0.002)***	0.483 (0.025)***
$\sigma_{\text{malt liquor}}$	-0.415 (0.004)***	0.260 (0.026)***
$\sigma_{\text{cocktails}}$	-0.616 (0.008)***	0.197 (0.039)***
σ_{flavored}	-1.915 (0.020)***	-0.866 (0.087)***
$\sigma_{\text{consumer packaged}}$	-1.744 (0.021)***	-1.328 (0.085)***
$\sigma_{\text{combo pack}}$	-0.471 (0.004)***	0.189 (0.028)***
Mean Own-price Elasticity	-6.878 (7.354)	-2.778 (0.904)
Mean Cross-price Elasticity	0.004 (0.022)	0.001 (0.011)
Market Fixed Effects	✓	✓
IV Type		gh_1, gh_2
Number of Products	2806	
Number of Observations	64497	

Bud Light boycott: Comparing Data-driven vs. Naive Nested Logit

Data-driven Nested Logit "Beer Light" category predicts better out-of-sample substitution

	Data-driven NL				Naive NL				Combined	
t=2023	-0.284***	(0.013)	-0.353***	(0.014)	-0.253***	(0.012)	-0.353***	(0.014)	-0.285***	(0.013)
Treatment same group	0.259***	(0.032)								
Regular beer × t=2023			0.087***	(0.025)						
Light beer × t=2023			0.327***	(0.033)						
Non-lager/craft × t=2023			0.024	(0.024)						
Boutique craft × t=2023			0.131***	(0.032)						
Naive light beer × t=2023					0.076**	(0.036)	0.176***	(0.036)		
Naive ale × t=2023							0.047**	(0.023)		
Naive regular beer × t=2023							0.198***	(0.024)		
Naive stout × t=2023							-0.028	(0.046)		
Only Data-driven Light Beer									0.252***	(0.037)
Only Naive Light Beer									0.010	(0.044)
Both Light Beer									0.280***	(0.057)
Observations	214353		209209		214353		209209		214353	
UPC FE	YES		YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES		YES	
Baseline group	Outside option		Outside option		Outside option		Outside option		Outside option	

*** p<0.01, ** p<0.05, * p<0.1

Mixed logit

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

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

where

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

and μ_{ijm} is drawn from

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

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

Mixed Logit

Structural estimates

Panel A: Linear Coefficients	β
log price	-3.790*** (0.091)
log unit size	3.086*** (0.063)
log unit quantity	3.232*** (0.063)
Panel B: Non-linear Coefficient	σ_p
log price	0.002*** (0.0003)
Own-price elasticity	-4.702 (2.875)
Cross-price elasticity	0.002 (0.005)
Market Fixed Effects	✓
IV Type	Gandhi-Houde
Number of Products	2806
Number of Observations	64497

Bud Light boycott comparison

	Data-driven Nested Logit				Naive Nested Logit				Mixed Logit			
t=2023	-0.284***	(0.013)	-0.353***	(0.014)	-0.253***	(0.012)	-0.353***	(0.014)	-0.255***	(0.032)	-0.232***	(0.015)
Treatment same group	0.259***	(0.032)										
Regular beer × t=2023			0.087***	(0.025)								
Light beer × t=2023			0.327***	(0.033)								
Non-lager/craft × t=2023			0.024	(0.024)								
Boutique craft × t=2023			0.131***	(0.032)								
Naive light beer × t=2023					0.076**	(0.036)	0.176***	(0.036)				
Naive ale × t=2023							0.047**	(0.023)				
Naive regular beer × t=2023							0.198***	(0.024)				
Naive stout × t=2023							-0.028	(0.046)				
Above median $\mathcal{E}^i_{\text{Bud Light}}$ × t=2023									0.013	(0.034)		
$\mathcal{E}^i_{\text{Bud Light}}$ × t=2023											-4.315	(3.240)
Observations	214353		209209		214353		209209		214353		214353	
UPC FE	YES		YES		YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES		YES		YES	
Baseline group	Outside option		Outside option		Outside option		Outside option		Below median		Outside option	

*** p<0.01, ** p<0.05, * p<0.1

Conclusions

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 with **minimal data and computational requirements**
- Nests correlated with political preferences
  $\longrightarrow$ Data-driven Nested Logit can capture unobservable factors of demand
  $\longrightarrow$ To the best of our knowledge, no empirical IO paper considers politics as a key demand driver
- For this application, our method outperforms Naive Nested Logit and Mixed Logit in out-of-sample test

Thanks!

Appendix

Blue-bus/Red-bus problem

Implications of IIA property

A traveler has a choice of commuting by car or taking a blue bus

Blue-bus/Red-bus problem

Implications of IIA property

A traveler has a choice of commuting by car or taking a blue bus

Assume indirect utility from *car* and *blue bus* is the same so

$$P_{car} = P_{blue\ bus} = \frac{1}{2} \implies \frac{P_{car}}{P_{blue\ bus}} = 1$$

Blue-bus/Red-bus problem

Implications of IIA property

A traveler has a choice of commuting by car or taking a blue bus

Assume indirect utility from *car* and *blue bus* is the same so

$$P_{car} = P_{blue\ bus} = \frac{1}{2} \implies \frac{P_{car}}{P_{blue\ bus}} = 1$$

Now a red bus is introduced, exactly equal to blue bus (but the color).

Blue-bus/Red-bus problem

Implications of IIA property

A traveler has a choice of commuting by car or taking a blue bus

Assume indirect utility from *car* and *blue bus* is the same so

$$P_{car} = P_{blue\ bus} = \frac{1}{2} \implies \frac{P_{car}}{P_{blue\ bus}} = 1$$

Now a red bus is introduced, exactly equal to blue bus (but the color). This implies:

$$\frac{P_{red\ bus}}{P_{blue\ bus}} = 1$$

Blue-bus/Red-bus problem

Implications of IIA property

A traveler has a choice of commuting by car or taking a blue bus

Assume indirect utility from *car* and *blue bus* is the same so

$$P_{car} = P_{blue\ bus} = \frac{1}{2} \implies \frac{P_{car}}{P_{blue\ bus}} = 1$$

Now a red bus is introduced, exactly equal to blue bus (but the color). This implies:

$$\frac{P_{red\ bus}}{P_{blue\ bus}} = 1$$

Given IIA, we still have $\frac{P_{car}}{P_{blue\ bus}} = 1$. The only consistent model with both is

$$P_{car} = P_{blue\ bus} = P_{red\ bus} = \frac{1}{3}$$

Connection to IIA

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

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 $\lambda_m^{k(j')}$

One-step group estimation

Combine steps 1 + 2 by solving the following constrained problem:

$$\underset{\substack{k(1),\dots,k(J)\\ \beta^1,\dots,\beta^K,\lambda_1^1,\dots,\lambda_M^K\\ \beta,\sigma^1,\dots,\sigma^K}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\left(\log\left(\frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}\right) - (x_{jm}\beta^k + \lambda_m^k)\right)^2,$$

where

$$\beta^k = \frac{\beta}{\sigma^k} \quad \text{and} \quad \lambda_m^k = (\sigma^k - 1)\log\left(\sum_{d\in B_k} e^{\log\left(\frac{\mathbb{P}_{dm}}{\mathbb{P}_{0m}}\right)-\lambda_m^k}\right),$$

One-step group estimation

Combine steps 1 + 2 by solving the following constrained problem:

$$\underset{\substack{k(1),\ldots,k(J)\\ \beta^1,\ldots,\beta^K,\lambda_1^1,\ldots,\lambda_M^K\\ \beta,\sigma^1,\ldots,\sigma^K}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\Big(\log\Big(\frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}\Big) - (x_{jm}\beta^k + \lambda_m^k)\Big)^2,$$

where

$$\beta^k = \frac{\beta}{\sigma^k} \quad \text{and} \quad \lambda_m^k = (\sigma^k - 1)\log\Big(\sum_{d\in B_k} e^{\log\left(\frac{\mathbb{P}_{dm}}{\mathbb{P}_{0m}}\right)-\lambda_m^k}\Big),$$

but

One-step group estimation

Combine steps 1 + 2 by solving the following constrained problem:

$$\underset{\substack{k(1),\dots,k(J)\\ \beta^1,\dots,\beta^K,\lambda_1^1,\dots,\lambda_M^K\\ \beta,\sigma^1,\dots,\sigma^K}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\left(\log\left(\frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}\right) - (x_{jm}\beta^k + \lambda_m^k)\right)^2,$$

where

$$\beta^k = \frac{\beta}{\sigma^k} \quad \text{and} \quad \lambda_m^k = (\sigma^k - 1)\log\left(\sum_{d\in B_k} e^{\log\left(\frac{\mathbb{P}_{dm}}{\mathbb{P}_{0m}}\right) - \lambda_m^k}\right),$$

but

- No theory has been developed for consistency of non-linear models
- In practice it may work, but substantial computational burden due to non-linearities

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 $M \to \infty$?
- Variation in x_{jm}, z_{jm} conditional on groups

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 $M \to \infty$?
- Variation in x_{jm}, z_{jm} conditional on groups
- Exogeneity of ξ_{jm}:

$$\mathbb{E}[\xi_{jm}, v_{jm} | x_{jm}, z_{jm}] = 0$$

- Thin tails
- $\lambda_m^{k(j)}$ 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 $k \neq k'$

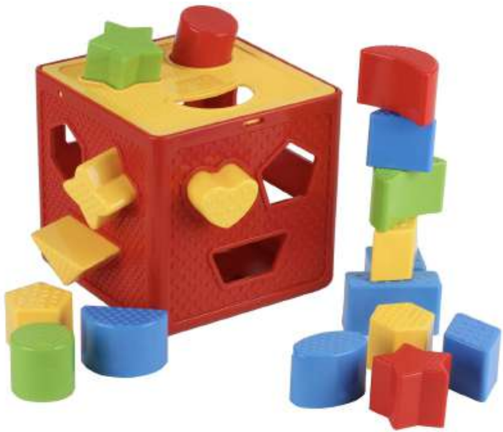
Regularity conditions with unknown nests

Regularity conditions to ensure consistent classification:

- K fixed. Let $J \to \infty$, $M \to \infty$, and $\frac{\log M}{J} \to 0$ Why $M \to \infty$?
- Variation in x_{jm}, z_{jm} conditional on groups
- Exogeneity of ξ_{jm}:
$$\mathbb{E}[\xi_{jm}, v_{jm} | x_{jm}, z_{jm}] = 0$$
- Thin tails
- $\lambda_m^{k(j)}$ 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 $k \neq k'$

Why asymptotics on M?

Graphical (literal) 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 $G = 2$:

$$y_{jm} = \alpha^*_{k^*_j} + \xi_{jm}, \qquad k_j \in \{1, 2\}.$$

- We characterize the misclassification probability:

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

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

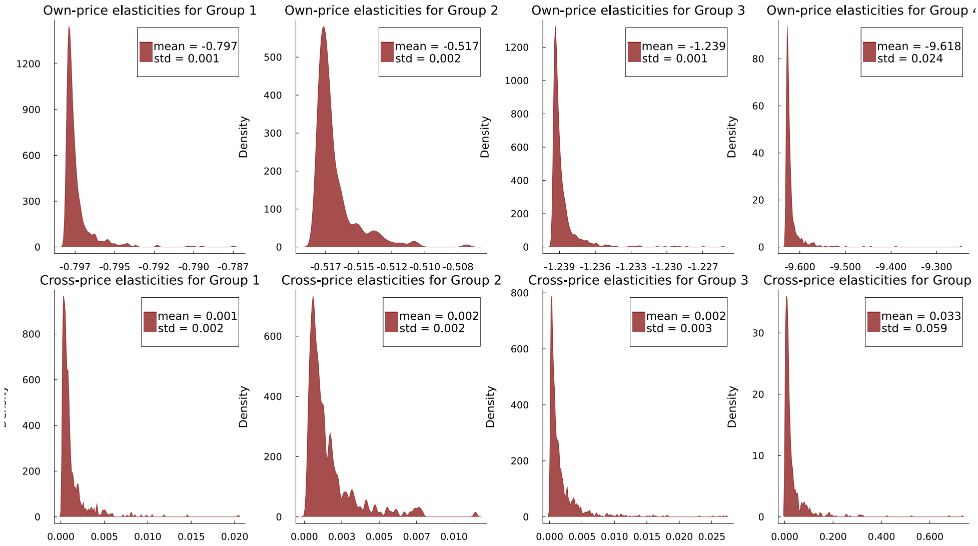
$$\Pr\left(\overline{\xi}_j > \frac{\alpha_1 + \alpha_2}{2} - \alpha^*_1\right) \quad = \quad 1 - \Phi\left(\frac{\sqrt{M}}{\sigma}\left(\frac{\alpha_1 + \alpha_2}{2} - \alpha^*_1\right)\right),$$

  which vanishes exponentially fast as M increases.

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

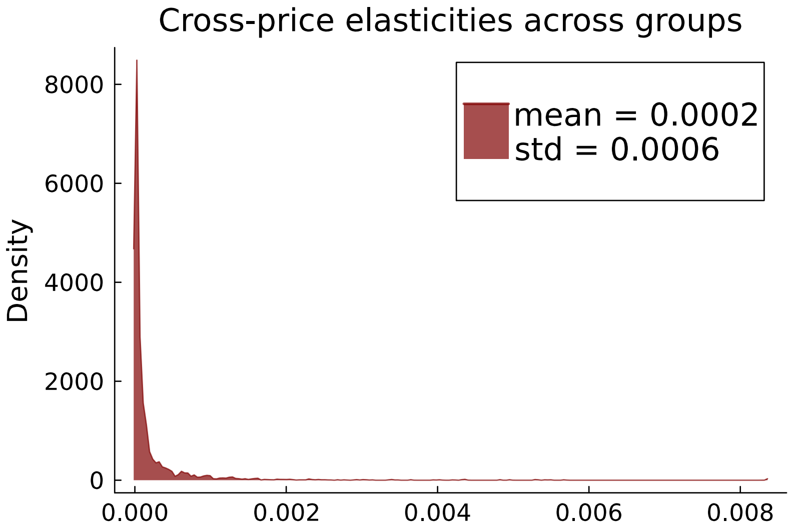
Substitution patterns: Within-group price elasticities

Back



Substitution Patterns: Cross-group Cross-price Elasticities

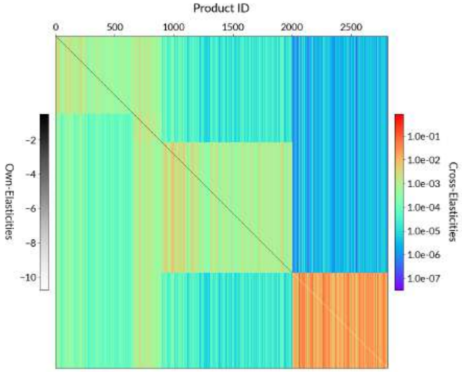
Back



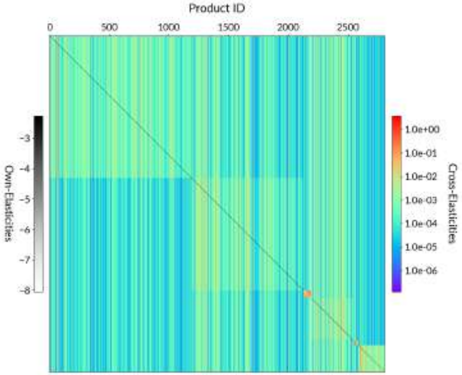
Comparison: Matrix of price elasticities

Data-driven NL captures more heterogeneity, while ML resembles Logit patterns

Back



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

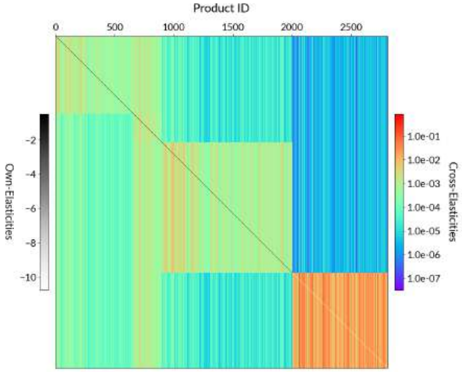


(b) Naive Nested Logit

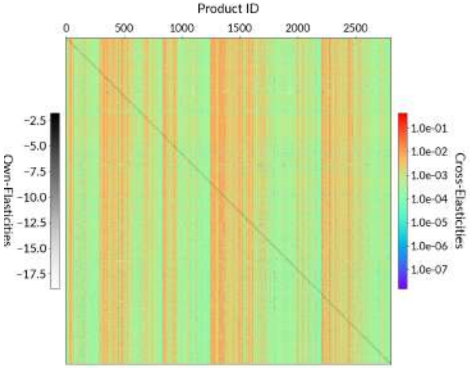
Comparison: Matrix of price elasticities

Data-driven NL captures more heterogeneity, while ML resembles Logit patterns

Back



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



(b) Mixed Logit

Mixed Logit

Bud Light boycott DiD results

Back

$$\log \mathbb{P}_{jmt} = \kappa_j + \eta_t + f(\text{elasticity}^j_{BudLight})\mathbb{1}\{t = 2023\} + \epsilon_{jmt}$$

	Continuous		DiD					
$\mathbb{1}\{t = 2023\}$	-0.232***	(0.015)	-0.224***	(0.016)	-0.247***	(0.013)	-0.353***	(0.014)
ML elasticity $\mathbb{1}\{t = 2023\}$	-4.315	(3.240)						
ML elasticity above median $\mathbb{1}\{t = 2023\}$			-0.041*	(0.023)				
ML elasticity top quartile $\mathbb{1}\{t = 2023\}$					0.013	(0.027)		
ML elasticity quartile 1 $\mathbb{1}\{t = 2023\}$							0.181***	(0.027)
ML elasticity quartile 2 $\mathbb{1}\{t = 2023\}$							0.078***	(0.027)
ML elasticity quartile 3 $\mathbb{1}\{t = 2023\}$							0.058**	(0.027)
ML elasticity quartile 4 $\mathbb{1}\{t = 2023\}$							0.119***	(0.027)
Observations	120265		213959		213959		208829	
UPC FE	YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES	
Baseline group	Outside option		Below median		Other quartiles		Outside option	

*** p<0.01, ** p<0.05, * p<0.1