Data-Driven Nests

Milena Almagro
(Chicago Booth & NBER)

Kenneth Lai
(LMU Munich)

Elena Manresa
(Princeton University)

Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

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: relaxes IIA but keeps closed form

Nested demand models are ubiquitous across many fields

- Industrial Organization: Verboven (1996), Parry & Small (2005), Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Berry et al. (2016), Björnerstedt & Verboven (2016), Allcott et al. (2024)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Verboven (1996), Parry & Small (2005), Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Berry et al. (2016), Björnerstedt & Verboven (2016), Allcott et al. (2024)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Verhoogen (2008), Atkeson & Burstein (2008), Khandelwal (2010), Fajgelbaum et al. (2011)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Verboven (1996), Parry & Small (2005), Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Berry et al. (2016), Björnerstedt & Verboven (2016), Allcott et al. (2024)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Verhoogen (2008), Atkeson & Burstein (2008), Khandelwal (2010), Fajgelbaum et al. (2011)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Verboven (1996), Parry & Small (2005), Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Berry et al. (2016), Björnerstedt & Verboven (2016), Allcott et al. (2024)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Verhoogen (2008), Atkeson & Burstein (2008), Khandelwal (2010), Fajgelbaum et al. (2011)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024)...
- Labor: Lamadon et al. (2022), Azar et al. (2022), Berger et al. (2022), Deb et al. (2024), Fogel et al. (ongoing)...

Nested demand models are ubiquitous across many fields

- Industrial Organization: Verboven (1996), Parry & Small (2005), Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Berry et al. (2016), Björnerstedt & Verboven (2016), Allcott et al. (2024)...
- Trade: Goldberg (1995), Broda & Weinstein (2006), Verhoogen (2008), Atkeson & Burstein (2008), Khandelwal (2010), Fajgelbaum et al. (2011)...
- Spatial Economics: Davis et al. (2019), Couture et al. (2023), Bordeu (2024)...
- Labor: Lamadon et al. (2022), Azar et al. (2022), Berger et al. (2022), Deb et al. (2024), Fogel et al. (ongoing)...

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

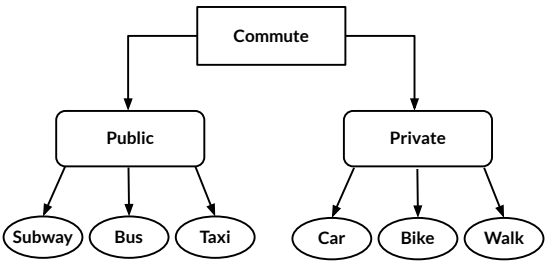
What's the right nesting structure?

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*.

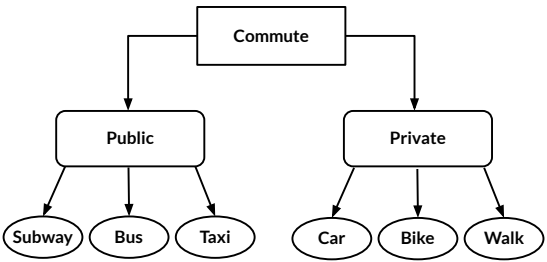
McFadden (1978)



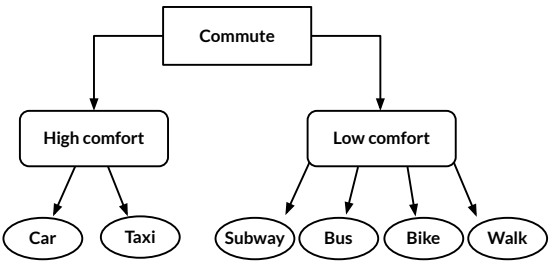
What's the right nesting structure?

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

McFadden (1978)



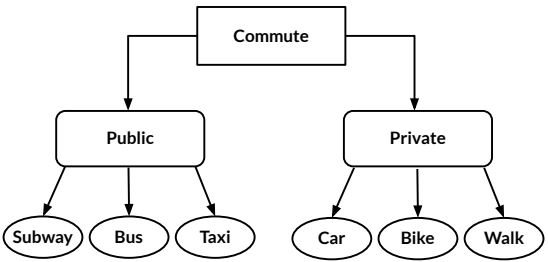
Alternative McFadden (1978)



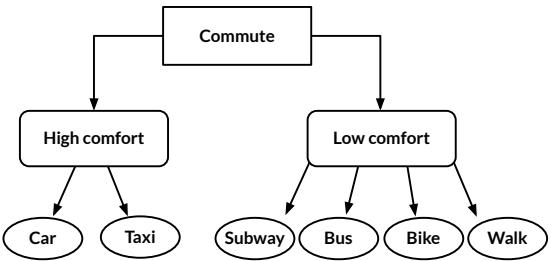
What's the right nesting structure?

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

McFadden (1978)



Alternative McFadden (1978)



$\Longrightarrow$ Mis-specification of nests can lead to biased estimates (Kovach & Tserenjigmid (2022))

This paper: Estimating nests in discrete choice models

- Identification relies on **exploiting the structure of the model**
- Two-step estimation procedure:
  1. Cluster products based on **model-implied demand curves** (Bonhomme and Manresa, 2015)
     → Demand curves share coefficients **within nests** but differ **across nests**
  2. Estimate model parameters as if the groups were known (Berry, 1994)

This paper: Estimating nests in discrete choice models

- Identification relies on **exploiting the structure of the model**
- Two-step estimation procedure:
  1. Cluster products based on **model-implied demand curves** (Bonhomme and Manresa, 2015)
     → Demand curves share coefficients **within nests** but differ **across nests**
  2. Estimate model parameters as if the groups were known (Berry, 1994)
- Panel data on aggregate market shares and product characteristics across markets

This paper: Estimating nests in discrete choice models

- Identification relies on **exploiting the structure of the model**
- Two-step estimation procedure:
  1. Cluster products based on **model-implied demand curves** (Bonhomme and Manresa, 2015)
     → Demand curves share coefficients **within nests** but differ **across nests**
  2. Estimate model parameters as if the groups were known (Berry, 1994)
- Panel data on aggregate market shares and product characteristics across markets
  - Our statistical justification exploits that the number of products and markets grows large

This paper: Estimating nests in discrete choice models

- Identification relies on **exploiting the structure of the model**
- Two-step estimation procedure:
  1. Cluster products based on **model-implied demand curves** (Bonhomme and Manresa, 2015)
     → Demand curves share coefficients **within nests** but differ **across nests**
  2. Estimate model parameters as if the groups were known (Berry, 1994)
- Panel data on aggregate market shares and product characteristics across markets
  - Our statistical justification exploits that the number of products and markets grows large
  - Under regularity conditions we prove:
    - The probability of misclassifying any product goes to zero
    - Nests are recovered at a faster-than-polynomial rate (**super-consistency**)
      → second step behaves as if nests were known

This paper: Estimating nests in discrete choice models

- Identification relies on **exploiting the structure of the model**
- Two-step estimation procedure:
  1. Cluster products based on **model-implied demand curves** (Bonhomme and Manresa, 2015)
     → Demand curves share coefficients **within nests** but differ **across nests**
  2. Estimate model parameters as if the groups were known (Berry, 1994)
- Panel data on aggregate market shares and product characteristics across markets
  - Our statistical justification exploits that the number of products and markets grows large
  - Under regularity conditions we prove:
    - The probability of misclassifying any product goes to zero
    - Nests are recovered at a faster-than-polynomial rate (**super-consistency**)
      → second step behaves as if nests were known
- Empirical application: Demand for beer
  - Data-driven NL outperforms Naive NL and Mixed Logit in an out-of-sample test

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), Fosgerau, Monardo & De Palma (2024)
- **Group fixed effect estimator:** Han & Moon (2010), Bonhomme & Manresa (2015), Choi and Okui (2024)
- **Eliciting grouping structure:** Kovach & Tserenjigmid (2022), Hortaçsu, Lieber, Monardo & de Paula (ongoing)

Our contribution: Recover the nesting structure from ***aggregate demand data*** for settings with ***many products*** exploiting the structure of the model

Outline

1. Empirical model
2. Practical aspects and extensions
3. Consistency, statistical properties, and Monte Carlo
4. Empirical application: Demand for beer

Empirical model

Empirical Model: Nested Logit

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

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

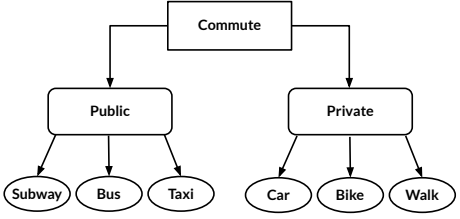
Empirical Model: Nested Logit

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

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

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, then

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



Empirical Model: Nested Logit

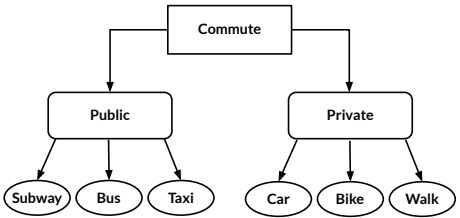
Consider the random utility model for agent i when choosing j in market m:

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

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, 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
  $\rightarrow \varepsilon_{ijm} = \eta_{i,k(j),m} + \tilde{\varepsilon}_{ijm}$ (Cardell, 1997)



Empirical Model: Nested Logit

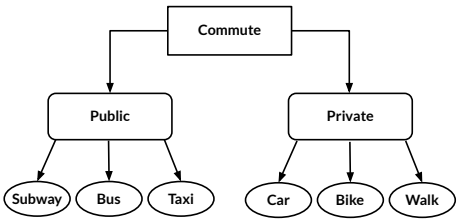
Consider the random utility model for agent i when choosing j in market m:

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

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, 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
  → $\varepsilon_{ijm} = \eta_{i,k(j),m} + \tilde{\varepsilon}_{ijm}$ (Cardell, 1997)
- $\sigma_{k(j)} \in (0, 1]$ captures within-nest independence
  → if $\sigma_k \to 0 \implies \eta_{k(j),m}$ dominates
  → if $\sigma_k = 1 \implies$ back to logit



Empirical Model: Nested Logit

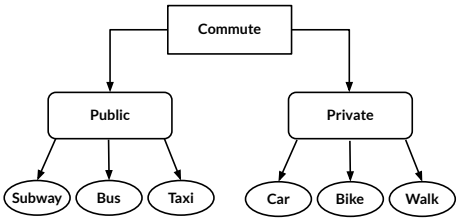
Consider the random utility model for agent i when choosing j in market m:

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

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, 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
  → $\varepsilon_{ijm} = \eta_{i,k(j),m} + \tilde{\varepsilon}_{ijm}$ (Cardell, 1997)
- $\sigma_{k(j)} \in (0, 1]$ captures within-nest independence
  → if $\sigma_k \to 0 \implies \eta_{k(j),m}$ dominates
  → if $\sigma_k = 1 \implies$ back to logit
- Nests lead to different degrees of substitution:
  → More substitution **within** than **across** nests



Towards an empirical equation

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, then

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

Towards an empirical equation

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, then

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

Towards an empirical equation

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, then

$$\mathbb{P}_{jm} = e^{\frac{\delta_{jm}}{\sigma_{k(j)}}} \frac{\left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma_{k(j)}}}\right)^{\sigma_{k(j)}-1}}{\sum_{l=1}^{K}\left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma_l}}\right)^{\sigma_l}} = e^{\frac{\delta_{jm}}{\sigma_{k(j)}}} \Lambda_{k(j)m} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma_{k(j)}} + \underbrace{\lambda_{k(j)m}}_{\log \Lambda_{k(j)m}}$$

Towards an empirical equation

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, then

$$\mathbb{P}_{jm} = e^{\frac{\delta_{jm}}{\sigma_{k(j)}}} \frac{\left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma_{k(j)}}}\right)^{\sigma_{k(j)}-1}}{\sum_{l=1}^{K}\left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma_l}}\right)^{\sigma_l}} = e^{\frac{\delta_{jm}}{\sigma_{k(j)}}} \Lambda_{k(j)m} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma_{k(j)}} + \underbrace{\lambda_{k(j)m}}_{\log \Lambda_{k(j)m}}$$

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

Towards an empirical equation

If $\varepsilon_{im} \overset{\text{i.i.d.}}{\sim}$ Nested Logit, then

$$\mathbb{P}_{jm} = e^{\frac{\delta_{jm}}{\sigma_{k(j)}}} \frac{\left(\sum_{d \in B_{k(j)}} e^{\frac{\delta_{dm}}{\sigma_{k(j)}}}\right)^{\sigma_{k(j)}-1}}{\sum_{l=1}^{K}\left(\sum_{d \in B_l} e^{\frac{\delta_{dm}}{\sigma_l}}\right)^{\sigma_l}} = e^{\frac{\delta_{jm}}{\sigma_{k(j)}}} \Lambda_{k(j)m} \implies \log \mathbb{P}_{jm} = \frac{\delta_{jm}}{\sigma_{k(j)}} + \underbrace{\lambda_{k(j)m}}_{\log \Lambda_{k(j)m}}$$

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

With a slight abuse of notation for ξ_{jm}, it follows:

$$\log \mathbb{P}_{jm} = \frac{\beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}}{\sigma_{k(j)}} + \lambda_{k(j)m} = \underbrace{\beta_{k(j),x}}_{\beta_x/\sigma_k} x_{jm} + \underbrace{\beta_{k(j),p}}_{\beta_p/\sigma_k} p_{jm} + \lambda_{k(j)m} + \xi_{jm}.$$

Outline

1. Empirical model
    1.1 Identification
    1.2 Estimation

2. Practical aspects and extensions

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

Grouping on model-implied demand curves

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}.$$

Grouping on model-implied demand curves

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}.$$

Key observation: $(\beta_{k(j),x},\ \beta_{k(j),p},\ \lambda_{k(j)m})$ are common for products in the same nest!

$\Longrightarrow$ **different demand curves** for products in different nests:

Grouping on model-implied demand curves

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}.$$

Key observation: $(\beta_{k(j),x},\ \beta_{k(j),p},\ \lambda_{k(j)m})$ are common for products in the same nest!

$\Longrightarrow$ **different demand curves** for products in different nests:

- $(\beta_{k(j),x},\ \beta_{k(j),p})$ are **nest-specific** slopes capturing sensitivity to covariates for nest k
  $\rightarrow$ With different σ_k, marginal effect of covariates differ across nests $\Longrightarrow$ $\beta_k \propto$ elasticities
- $\lambda_{k(j)m}$ is a **nest-by-market** fixed effect capturing within-market **nest popularity**
  $\rightarrow$ Example: craft beer is more popular in the East and West Coasts relative to the South

Grouping on model-implied demand curves

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}.$$

Key observation: $(\beta_{k(j),x},\ \beta_{k(j),p},\ \lambda_{k(j)m})$ are common for products in the same nest!

$\Longrightarrow$ **different demand curves** for products in different nests:

- $(\beta_{k(j),x},\ \beta_{k(j),p})$ are **nest-specific** slopes capturing sensitivity to covariates for nest k
  $\rightarrow$ With different σ_k, marginal effect of covariates differ across nests $\Longrightarrow$ $\beta_k \propto$ elasticities
- $\lambda_{k(j)m}$ is a **nest-by-market** fixed effect capturing within-market **nest popularity**
  $\rightarrow$ Example: craft beer is more popular in the East and West Coasts relative to the South

Identification requires:

- **Group separation:** Cannot have $\beta_k = \beta_{k'}$ AND $\lambda_{km} = \lambda_{k'm},\ \forall m$ for some $k \neq k'$
- **No grouping structure** $e(j)$ **in the error** ξ_{jm}: only allowed if $B_k \subseteq B_e$

Intuition on nest identification

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}$$

Intuition on nest identification

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}$$

Prices are endogenous. Replace p_{jm} by $\hat{p}_{jm}$, the price purged of endogeneity (details later):

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} \hat{p}_{jm} + \lambda_{k(j)m} + \xi_{jm}, \quad \mathbb{E}[\xi_{jm} \mid x_{jm}, \hat{p}_{jm}] = 0$$

Intuition on nest identification

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}$$

Prices are endogenous. Replace p_{jm} by $\hat{p}_{jm}$, the price purged of endogeneity (details later):

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} \hat{p}_{jm} + \lambda_{k(j)m} + \xi_{jm}, \quad \mathbb{E}[\xi_{jm} \mid x_{jm}, \hat{p}_{jm}] = 0$$

Key observation: Data are explained by K linear fits, one per nest, with coefficients $(\beta_{k,x}, \beta_{k,p}, \lambda_{km})$

Intuition on nest identification

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}$$

Prices are endogenous. Replace p_{jm} by $\hat{p}_{jm}$, the price purged of endogeneity (details later):

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} \hat{p}_{jm} + \lambda_{k(j)m} + \xi_{jm}, \quad \mathbb{E}[\xi_{jm} \mid x_{jm}, \hat{p}_{jm}] = 0$$

Key observation: Data are explained by K linear fits, one per nest, with coefficients $(\beta_{k,x}, \beta_{k,p}, \lambda_{km})$

Find the partition of products that best explains the data with K linear fits.

Simultaneously search over:

1. Coefficients: $(\beta_1, \ldots, \beta_K, \lambda_{11}, \ldots, \lambda_{KM})$

Intuition on nest identification

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} p_{jm} + \lambda_{k(j)m} + \xi_{jm}$$

Prices are endogenous. Replace p_{jm} by $\hat{p}_{jm}$, the price purged of endogeneity (details later):

$$\log \mathbb{P}_{jm} = \beta_{k(j),x} x_{jm} + \beta_{k(j),p} \hat{p}_{jm} + \lambda_{k(j)m} + \xi_{jm}, \quad \mathbb{E}[\xi_{jm} \mid x_{jm}, \hat{p}_{jm}] = 0$$

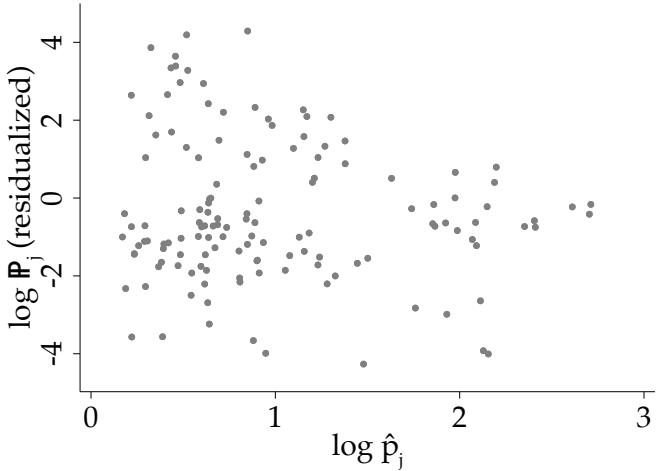
Key observation: Data are explained by K linear fits, one per nest, with coefficients $(\beta_{k,x}, \beta_{k,p}, \lambda_{km})$

Find the partition of products that best explains the data with K linear fits.

Simultaneously search over:

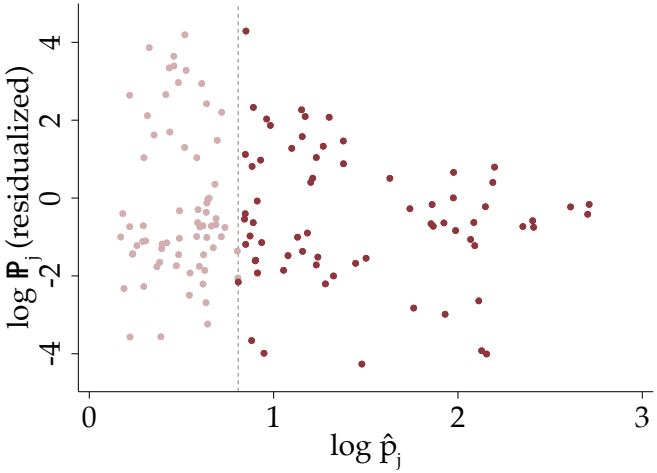
1. Coefficients: $(\beta_1, \ldots, \beta_K, \lambda_{11}, \ldots, \lambda_{KM})$
2. Partitions: $\{k(j)\}_{j=1}^{J}$ to assign products to different lines

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



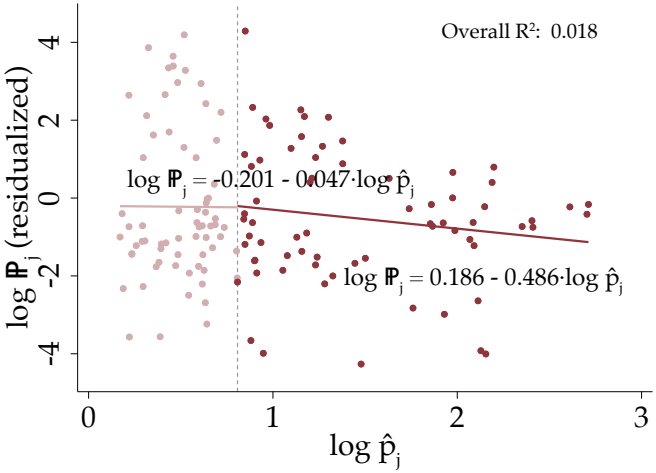
Note: NielsenIQ beer data. Each dot is a product, averaged across markets.

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



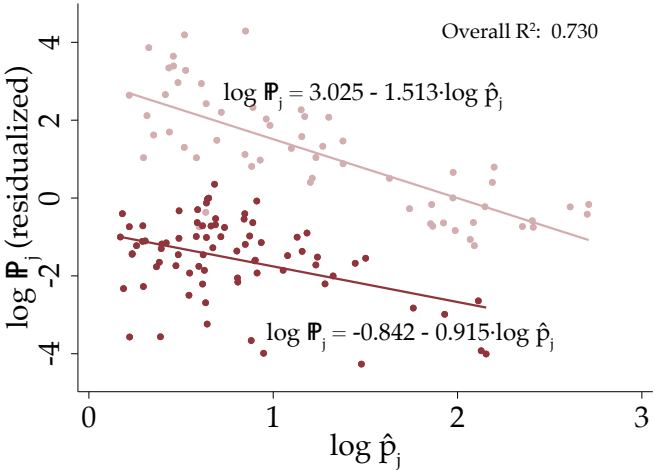
Note: NielsenIQ beer data. Each dot is a product, averaged across markets.

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



Note: NielsenIQ beer data. Each dot is a product, averaged across markets.

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



Note: NielsenIQ beer data. Each dot is a product, averaged across markets.

Outline

1. Empirical model
   - 1.1 Identification
   - 1.2 Estimation

2. Practical aspects and extensions

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

Estimation of nests

First step: Classification

Take $\hat{p}_{jm}$ prices purged of endogeneity (projection of p_{jm} on x_{jm} and a price shifter z_{jm})

Estimation of nests

First step: Classification

Take $\hat{p}_{jm}$ prices purged of endogeneity (projection of p_{jm} on x_{jm} and a price shifter z_{jm})

Grouped fixed-effects estimator defined by the following optimization problem:

$$\underset{\substack{k(1),\ldots,k(J)\\ \beta_1,\ldots,\beta_K,\lambda_{11},\ldots,\lambda_{KM}}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\left(\log \mathbb{P}_{jm} - (\beta_{k(j),p}\hat{p}_{jm} + \beta_{k(j),x}x_{jm} + \lambda_{k(j)m})\right)^2$$

Extends k-means by minimizing distance to regression lines (instead of cluster centroids):

- If nests were known, this is just a regression with nest interactions
- If we exclude covariates, $\hat{p}_{jm}$ and x_{jm}, we're back to k-means

Estimation of nests

First step: Classification

Take $\hat{p}_{jm}$ prices purged of endogeneity (projection of p_{jm} on x_{jm} and a price shifter z_{jm})

Grouped fixed-effects estimator defined by the following optimization problem:

$$\underset{\substack{k(1),\ldots,k(J)\\ \beta_1,\ldots,\beta_K,\lambda_{11},\ldots,\lambda_{KM}}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\left(\log \mathbb{P}_{jm} - (\beta_{k(j),p}\hat{p}_{jm} + \beta_{k(j),x}x_{jm} + \lambda_{k(j)m})\right)^2$$

Extends k-means by minimizing distance to regression lines (instead of cluster centroids):

- If nests were known, this is just a regression with nest interactions
- If we exclude covariates, $\hat{p}_{jm}$ and x_{jm}, we're back to k-means

Combinatorial, non-convex problem!

→ Approach: iterative clustering algorithm extending Bonhomme and Manresa (2015)

Estimation of structural parameters

Second step

Once we have the nesting structure $\{k(1), \ldots, k(J)\}$, estimate β and σ_k as if nests were known

Estimation of structural parameters

Second step

Once we have the nesting structure $\{k(1), \ldots, k(J)\}$, estimate β and σ_k as if nests were known

We follow Berry (1994):

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

Estimation of structural parameters

Second step

Once we have the nesting structure $\{k(1), \ldots, k(J)\}$, estimate β and σ_k as if nests were known

We follow Berry (1994):

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

Address (potential) endogeneity with GMM estimation:

- Instrument for prices, z_{jm}
  → Note: Restrict to settings with relevant instruments as $J \to \infty$, Armstrong (2016) Nash-Bertrand example

Estimation of structural parameters

Second step

Once we have the nesting structure $\{k(1), \ldots, k(J)\}$, estimate β and σ_k as if nests were known

We follow Berry (1994):

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

Address (potential) endogeneity with GMM estimation:

- Instrument for prices, z_{jm}
  → Note: Restrict to settings with relevant instruments as $J \to \infty$, Armstrong (2016) Nash-Bertrand example
- Instrument for $\log \mathbb{P}_{j|k(j),m} = \frac{e^{\frac{\beta_p p_{jm} + \beta_x x_{jm}}{\sigma_{k(j)}}}}{\sum_{d \in B_{k(j)}} e^{\frac{\beta_p p_{dm} + \beta_x x_{dm}}{\sigma_{k(j)}}}}$ $\implies$ Functions of $x_{j'm}$ for $j' \in B_{k(j)}$

Practical aspects and extensions

Outline

1. Empirical model

2. Practical aspects and extensions
   - 2.1 Endogenous prices
   - 2.2 Choosing the number of nests
   - 2.3 Introducing consumer heterogeneity
   - 2.4 Higher-level nesting structures
   - 2.5 Type II EV error: Nested CES

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

Dealing with price endogeneity

When prices are endogenous, $\mathbb{E}[\xi_{jm} \mid x_{jm}, p_{jm}] \neq 0$, need an excluded shifter z_{jm} with $\mathbb{E}[\xi_{jm} \mid x_{jm}, z_{jm}] = 0$

Dealing with price endogeneity

When prices are endogenous, $\mathbb{E}[\xi_{jm} \mid x_{jm}, p_{jm}] \neq 0$, need an excluded shifter z_{jm} with $\mathbb{E}[\xi_{jm} \mid x_{jm}, z_{jm}] = 0$

Project prices on exogenous variables with **group-specific slopes** and **group-market intercepts**:

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

estimated with the same classification algorithm, over price groups $k'(1), \ldots, k'(J)$

Dealing with price endogeneity

When prices are endogenous, $\mathbb{E}[\xi_{jm} \mid x_{jm}, p_{jm}] \neq 0$, need an excluded shifter z_{jm} with $\mathbb{E}[\xi_{jm} \mid x_{jm}, z_{jm}] = 0$

Project prices on exogenous variables with **group-specific slopes** and **group-market intercepts**:

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

estimated with the same classification algorithm, over price groups $k'(1), \ldots, k'(J)$

Why groups in the projection?

- First step must include every exogenous regressor of the demand equation
  → Demand has nest slopes and nest-market intercepts $\implies$ so must the projection
- Otherwise $\hat{p}_{jm}$ keeps group variation correlated with λ_{km} $\implies$ inconsistent second step
- Price groups $g(j)$ need not coincide with nests $k(j)$
  → e.g. multiproduct firms with common marginal costs across their portfolio

Why not GMM?

Outline

1. Empirical model

2. Practical aspects and extensions
   - 2.1 Endogenous prices
   - 2.2 Choosing the number of nests
   - 2.3 Introducing consumer heterogeneity
   - 2.4 Higher-level nesting structures
   - 2.5 Type II EV error: Nested CES

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

Choosing K: Cross-validation with Elbow method

We can estimate the number of nests using a **cross-validation** procedure

Choosing K: Cross-validation with Elbow method

We can estimate the number of nests using a **cross-validation** procedure

For $K \in \{1, ..., \bar{K}\}$:

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

$$MSE_n(K) = \frac{1}{|P_n| \cdot M} \sum_{m=1}^{M} \sum_{j \in P_n} \left(\log \mathbb{P}_{jm} - \beta_{k(j),x,-n} x_{jm} - \beta_{k(j),p,-n} \hat{p}_{jm} - \lambda_{k(j)m,-n} \right)^2$$

- Take average across N folds:

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

- Choose K^* at the elbow: where $MSE(K) - MSE(K+1)$ becomes small

Outline

1. Empirical model

2. Practical aspects and extensions
   - 2.1 Endogenous prices
   - 2.2 Choosing the number of nests
   - 2.3 Introducing consumer heterogeneity
   - 2.4 Higher-level nesting structures
   - 2.5 Type II EV error: Nested CES

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

Individual heterogeneity with observed conditional shares

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

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

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

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

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 $\mathbb{P}_{jm}^{t} = \mathbb{P}_{m}(j|\omega^{t})$ are observed, 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

1. Empirical model

2. Practical aspects and extensions
   - 2.1 Endogenous prices
   - 2.2 Choosing the number of nests
   - 2.3 Introducing consumer heterogeneity
   - 2.4 Higher-level nesting structures
   - 2.5 Type II EV error: Nested CES

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

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:

$$\log \mathbb{P}_{jm} = \beta_k x_{jm} + \left(\frac{\sigma_k}{\sigma_n} - 1\right) \log IV_{km} + (\sigma_n - 1) \log IV_{nm} - \log IV_m$$

so that

$$\log \mathbb{P}_{jm} = \beta_k x_{jm} + \lambda_{km} + \lambda_{n(k)m}$$

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:

$$\log \mathbb{P}_{jm} = \beta_k x_{jm} + \left(\frac{\sigma_k}{\sigma_n} - 1\right) \log IV_{km} + (\sigma_n - 1) \log IV_{nm} - \log IV_m$$

so that

$$\log \mathbb{P}_{jm} = \beta_k x_{jm} + \lambda_{km} + \lambda_{n(k)m}$$

We can run the same classification algorithm!
→ Since $n = n(k)$, $\lambda_{km} + \lambda_{n(k)m}$ is a nest-market intercept: the first step is identical to the baseline
→ The upper level only matters in the second step

Note: needs normalization of some $\lambda_{km} = 0$ to avoid co-linear group-market fixed effects.

Outline

1. Empirical model

2. Practical aspects and extensions
   - 2.1 Endogenous prices
   - 2.2 Choosing the number of nests
   - 2.3 Introducing consumer heterogeneity
   - 2.4 Higher-level nesting structures
   - 2.5 Type II EV error: Nested CES

3. Consistency, statistical properties, and Monte Carlo

4. Empirical application: Demand for beer

Type II EV error

Nested CES

A representative consumer maximizes a two-nested CES utility function over products j:

$$U_m = \left(\sum_{k=1}^{K} \left(\sum_{j \in B_k} \left(A_{jm}\, c_{jm} \right)^{(\sigma_k - 1)/\sigma_k} \right)^{\frac{\sigma_k}{\sigma_k - 1} \cdot \frac{\sigma_0 - 1}{\sigma_0}} \right)^{\sigma_0/(\sigma_0 - 1)},$$

where

- c_{jm} is consumption of product j,
- A_{jm} is a quality shifter,
- $\sigma_k > 1$ is the within-nest elasticity of substitution,
- $\sigma_0 \geq 1$ is the across-nest elasticity.

Type II EV error

Expenditure shares

Standard expenditure minimization gives the expenditure share of product j:

$$\mathbb{P}_{jm} = \frac{\left(A_{jm}/p_{jm}\right)^{\sigma_k - 1}}{\sum_{j' \in B_k} \left(A_{j'm}/p_{j'm}\right)^{\sigma_k - 1}} \cdot \frac{P_{km}^{1-\sigma_0}}{\sum_{k'=1}^{K} P_{k'm}^{1-\sigma_0}} = \left(\frac{A_{jm}}{p_{jm}}\right)^{\sigma_{k(j)} - 1} \Lambda_{k(j)m},$$

where $\Lambda_{k(j)m}$ collects all nest- and market-level terms.

Type II EV error

Expenditure shares

Standard expenditure minimization gives the expenditure share of product j:

$$\mathbb{P}_{jm} = \frac{\left(A_{jm}/p_{jm}\right)^{\sigma_k - 1}}{\sum_{j' \in B_k} \left(A_{j'm}/p_{j'm}\right)^{\sigma_k - 1}} \cdot \frac{P_{km}^{1-\sigma_0}}{\sum_{k'=1}^{K} P_{k'm}^{1-\sigma_0}} = \left(\frac{A_{jm}}{p_{jm}}\right)^{\sigma_{k(j)} - 1} \Lambda_{k(j)m},$$

where $\Lambda_{k(j)m}$ collects all nest- and market-level terms.

Taking logs and setting $\log A_{jm} = \beta_x x_{jm} + \xi_{jm}$ (with the same abuse of notation for ξ_{jm}):

$$\log \mathbb{P}_{jm} = \underbrace{(\sigma_{k(j)} - 1)\beta_x}_{\beta_{k(j),x}} x_{jm} - \underbrace{(\sigma_{k(j)} - 1)}_{-\beta_{k(j),p}} \log p_{jm} + \lambda_{k(j)m} + \xi_{jm},$$

which exactly maps to the first-step equation of our baseline model.

Consistency, statistical properties, and Monte Carlo

Conditions for consistent classification

1. **Sample size:** $J, M \to \infty$ with $J = O(M^{q_1})$ and $M = O(J^{q_2})$
   $\to$ Many products and many markets, each at most a power of the other

Conditions for consistent classification

1. **Sample size:** $J, M \to \infty$ with $J = O(M^{q_1})$ and $M = O(J^{q_2})$
   $\to$ Many products and many markets, each at most a power of the other
2. **Nests:** K fixed, $J_k/J \geq \underline{\pi} > 0$, and σ_k bounded away from 0 and 1
   $\to$ Every nest keeps a positive share of the market as J grows

Conditions for consistent classification

1. **Sample size:** $J, M \to \infty$ with $J = O(M^{q_1})$ and $M = O(J^{q_2})$
   → Many products and many markets, each at most a power of the other
2. **Nests:** K fixed, $J_k/J \geq \underline{\pi} > 0$, and σ_k bounded away from 0 and 1
   → Every nest keeps a positive share of the market as J grows
3. **Demand shocks:** Markets independent. $\mathbb{E}[\xi_{jm} \mid x_m, z_m] = 0$, ξ_{jm} weakly dependent across j
   → e.g. firms do not reposition products in response to rivals' shocks

Conditions for consistent classification

1. **Sample size:** $J, M \to \infty$ with $J = O(M^{q_1})$ and $M = O(J^{q_2})$
   → Many products and many markets, each at most a power of the other
2. **Nests:** K fixed, $J_k/J \geq \underline{\pi} > 0$, and σ_k bounded away from 0 and 1
   → Every nest keeps a positive share of the market as J grows
3. **Demand shocks:** Markets independent. $\mathbb{E}[\xi_{jm} \mid x_m, z_m] = 0$, ξ_{jm} weakly dependent across j
   → e.g. firms do not reposition products in response to rivals' shocks
4. **Prices:** $\mathbb{E}[\omega_{jm} x_{jm}] = \mathbb{E}[\omega_{jm} z_{jm}] = 0$, ω_{jm} weakly dependent across j, and $\gamma_z^g \neq 0$
   → Price shifter moves prices; price residual orthogonal to exogenous variables

Conditions for consistent classification

1. **Sample size:** $J, M \to \infty$ with $J = O(M^{q_1})$ and $M = O(J^{q_2})$
   → Many products and many markets, each at most a power of the other
2. **Nests:** K fixed, $J_k/J \geq \underline{\pi} > 0$, and σ_k bounded away from 0 and 1
   → Every nest keeps a positive share of the market as J grows
3. **Demand shocks:** Markets independent. $\mathbb{E}[\xi_{jm} \mid x_m, z_m] = 0$, ξ_{jm} weakly dependent across j
   → e.g. firms do not reposition products in response to rivals' shocks
4. **Prices:** $\mathbb{E}[\omega_{jm} x_{jm}] = \mathbb{E}[\omega_{jm} z_{jm}] = 0$, ω_{jm} weakly dependent across j, and $\gamma_z^g \neq 0$
   → Price shifter moves prices; price residual orthogonal to exogenous variables
5. **Separation:** $\frac{1}{M} \sum_m (\lambda_{km} - \lambda_{lm})^2 \geq c > 0$ for $k \neq l$; same for (γ^g, κ_{gm})
   → Nests differ in popularity across markets; price groups differ in pricing equations

Conditions for consistent classification

1. **Sample size:** $J, M \to \infty$ with $J = O(M^{q_1})$ and $M = O(J^{q_2})$
   → Many products and many markets, each at most a power of the other
2. **Nests:** K fixed, $J_k/J \geq \underline{\pi} > 0$, and σ_k bounded away from 0 and 1
   → Every nest keeps a positive share of the market as J grows
3. **Demand shocks:** Markets independent. $\mathbb{E}[\xi_{jm} \mid x_m, z_m] = 0$, ξ_{jm} weakly dependent across j
   → e.g. firms do not reposition products in response to rivals' shocks
4. **Prices:** $\mathbb{E}[\omega_{jm} x_{jm}] = \mathbb{E}[\omega_{jm} z_{jm}] = 0$, ω_{jm} weakly dependent across j, and $\gamma_z^g \neq 0$
   → Price shifter moves prices; price residual orthogonal to exogenous variables
5. **Separation:** $\frac{1}{M} \sum_m (\lambda_{km} - \lambda_{lm})^2 \geq c > 0$ for $k \neq l$; same for (γ^g, κ_{gm})
   → Nests differ in popularity across markets; price groups differ in pricing equations
6. **Within-nest variation:** $(x_{jm}, \hat{p}_{jm})$ vary across products within nest and market
   → No two products of a nest move together across all markets

Two main results

Theorem 1 (Consistent classification). Under the conditions above, for every $\varsigma > 0$,

$$\mathbb{P}(\hat{k}(j) \neq k(j) \text{ for some } j) = o(1) + o(JM^{-\varsigma})$$

- Per-product misclassification vanishes faster than any power of $M \implies$ super-consistency
- Nests are recovered even when $M \ll J$

Two main results

Theorem 1 (Consistent classification). Under the conditions above, for every $\varsigma > 0$,

$$\mathbb{P}(\hat{k}(j) \neq k(j) \text{ for some } j) = o(1) + o(JM^{-\varsigma})$$

- Per-product misclassification vanishes faster than any power of $M \implies$ super-consistency
- Nests are recovered even when $M \ll J$

Theorem 2 and Corollary 1 (Oracle equivalence). Let $\theta = (\beta, \sigma_1, \dots, \sigma_K)$. Let $\hat{\theta}(\hat{k})$ be the second-step estimator in levels with estimated nests and $\hat{\theta}(k)$ the infeasible one with true nests. Then

$$\mathbb{P}(\hat{\theta}(\hat{k}) \neq \hat{\theta}(k)) \to 0 \quad \text{and} \quad \sqrt{JM}\, D_J^{-1}(\hat{\theta}(\hat{k}) - \theta_0) \to_d N(0, V), \quad D_J = \text{diag}(I, (\log J)^{-1} I_K)$$

- β converges at rate $\sqrt{JM}$, σ_k at rate $\log J\sqrt{JM}$
- Same asymptotic distribution as if nests were known
  $\implies$ no adjustment of standard errors, no sample splitting

Discussion

We **cannot** accommodate:

- Unobservables ξ_{jm} with a grouping structure not spanned by the nests B_k
  $\rightarrow$ The classifier may pick up those groups instead of true nests
- Non-linear utility in parameters
  $\rightarrow$ No theory has been successfully developed for non-linear models (yet?)
  $\rightarrow$ Tackling endogeneity would require an alternative strategy, like a control function approach
  $\rightarrow$ But linearity (or log-linearity) in parameters is prevalent in empirical work

Discussion

We **cannot** accommodate:

- Unobservables ξ_{jm} with a grouping structure not spanned by the nests B_k
  $\rightarrow$ The classifier may pick up those groups instead of true nests
- Non-linear utility in parameters
  $\rightarrow$ No theory has been successfully developed for non-linear models (yet?)
  $\rightarrow$ Tackling endogeneity would require an alternative strategy, like a control function approach
  $\rightarrow$ But linearity (or log-linearity) in parameters is prevalent in empirical work

Our asymptotics require:

- J growing large
  - Required for consistency of the estimates $\hat{\lambda}_{km}$
  - But also to make sure that population λ_{km} are well defined finite objects
  - Fixed J requires strong assumptions (such as x_{jm}, p_{jm}, and ξ_{jm} bounded)

Discussion

We **cannot** accommodate:

- Unobservables ξ_{jm} with a grouping structure not spanned by the nests B_k
  → The classifier may pick up those groups instead of true nests
- Non-linear utility in parameters
  → No theory has been successfully developed for non-linear models (yet?)
  → Tackling endogeneity would require an alternative strategy, like a control function approach
  → But linearity (or log-linearity) in parameters is prevalent in empirical work

Our asymptotics require:

- J growing large
  - Required for consistency of the estimates $\hat{\lambda}_{km}$
  - But also to make sure that population λ_{km} are well defined finite objects
  - Fixed J requires strong assumptions (such as x_{jm}, p_{jm}, and ξ_{jm} bounded)
- M growing large:
  - In the presence of ξ_{jm}, this ensures probability of misclassification $\to 0$ as $M \to \infty$ Details

Monte Carlo design: Data generating process

Fix $K = 3$. Set $\boldsymbol{\sigma} = (0.2, 0.3, 0.6)$. Assign products to nests with uniform draws $k(j) \sim \mathcal{U}\{1, \ldots, K\}$

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

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

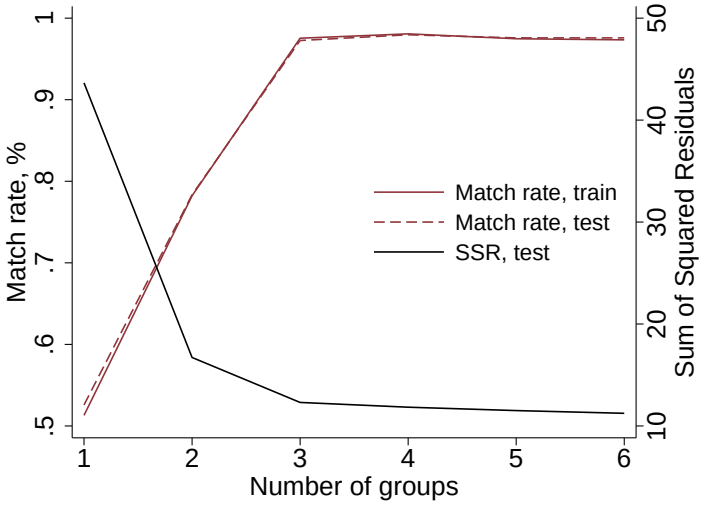
with $(\beta_p, \beta_x) = (-1, 1)$, $\rho = 0.3$, and mc_{jm} as the excluded price shifter z_{jm}.

Generate data as follows:

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

Estimating the number of nests: Monte Carlo

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



Results with $K = 3$

50 Monte Carlo replications

					β_p	β_c	σ_1	σ_2	σ_3
J	M	Runtime	Matched	True	-1	1	0.2	0.3	0.6
100	10	00:01	0.988	Mean β	-1.001	0.992	0.195	0.298	0.599
				Std β	0.037	0.031	0.013	0.007	0.01
100	50	00:18	1.0	Mean β	-1.0	1.003	0.2	0.3	0.6
				Std β	0.012	0.012	0.001	0.002	0.004
100	100	01:42	1.0	Mean β	-1.001	1.001	0.2	0.3	0.6
				Std β	0.008	0.008	0.001	0.001	0.003
500	10	00:14	0.995	Mean β	-1.0	0.998	0.199	0.3	0.6
				Std β	0.013	0.012	0.006	0.005	0.003
500	50	03:59	1.0	Mean β	-1.001	1.0	0.2	0.3	0.6
				Std β	0.006	0.004	0.001	0.001	0.001
500	100	17:42	1.0	Mean β	-1.0	1.0	0.2	0.3	0.6
				Std β	0.004	0.003	0.0	0.001	0.001
1000	10	00:15	0.995	Mean β	-1.001	1.002	0.2	0.3	0.599
				Std β	0.01	0.011	0.002	0.002	0.002
1000	50	08:08	1.0	Mean β	-1.002	1.001	0.2	0.3	0.6
				Std β	0.003	0.004	0.0	0.0	0.001
1000	100	31:48	1.0	Mean β	-1.0	1.0	0.2	0.3	0.6
				Std β	0.003	0.003	0.0	0.0	0.001

Notes: Running time for a 32-core machine with 16GB of RAM per core.

Empirical application: Demand for beer

Outline

1. Empirical model

2. Practical aspects and extensions

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 Out-of-sample group validation
   - 4.4 Second-step estimation
   - 4.5 Comparison with alternative empirical models

NielsenIQ 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 units sold by UPC and state, aggregating across stores
  - Characteristics: Prices, unit size, units per UPC, type of beer, brand, packaging, domestic

NielsenIQ 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 units sold by UPC and state, aggregating across stores
  - Characteristics: Prices, unit size, units per UPC, type of beer, brand, packaging, domestic

Keep products that are sold in at least ten states. Remaining UPCs form the outside option

NielsenIQ 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 units sold by UPC and state, aggregating across stores
  - Characteristics: Prices, unit size, units per UPC, type of beer, brand, packaging, domestic

Keep products that are sold in at least ten states. Remaining UPCs form the outside option

- Unbalanced panel with $J = 2,806$ unique UPCs (11.6%) and 64,497 observations
- Cover between 61.5% and 99% of volume sales across states
- Average UPC is sold across 23.32 states $\sim$ 50% missing markets
  $\rightarrow$ Algorithm adapted to entry and exit of products across markets

Summary statistics

Large variation in shares and covariates

Panel A: Full Sample Summary Statistics

Variable	Mean	Std. Dev.	10th pct	Median	90th pct
Price ($)	10.618	6.758	2.596	10.298	20.308
Unit quantity (units)	8.838	8.229	1.000	6.000	24.000
Price per ml ($)	0.004	0.001	0.002	0.004	0.005
Unit size (ml)	535.627	450.385	354.882	354.882	739.338
Market share	0.859%	1.152%	0.061%	0.498%	1.967%

Panel B: Median Characteristics of the top ten national brands

Brand	Market share	Unit price/ml	Unit price	Unit quantity	Unit size
Bud Light	12.53%	0.003	8.63	6.00	473.18
Miller	7.81%	0.003	12.47	12.00	354.88
Modelo	7.63%	0.005	10.23	6.00	354.88
Coors	7.11%	0.003	12.62	12.00	354.88
Bud	6.73%	0.003	3.22	1.00	739.34
Michelob	5.96%	0.004	15.20	12.00	354.88
Corona	4.59%	0.004	16.52	12.00	354.88
Busch	4.43%	0.002	10.87	12.00	354.88
Natural	3.63%	0.002	6.49	6.00	473.18
New Belgium	1.98%	0.005	10.28	1.00	567.81

Empirical Model

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

$$\delta_{jm} = \beta_p \log \text{price}_{jm} + \beta_s \log \text{size}_j + \beta_n \log \text{units}_j + \varphi_m + \xi_{jm}$$

where size_j is unit size, units_j is units per UPC, and φ_m is a market fixed effect
$\rightarrow$ We later consider an extension with categorical variables based on NielsenIQ beer types.

Empirical Model

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

$$\delta_{jm} = \beta_p \log \text{price}_{jm} + \beta_s \log \text{size}_j + \beta_n \log \text{units}_j + \varphi_m + \xi_{jm}$$

where size_j is unit size, units_j is units per UPC, and φ_m is a market fixed effect
$\rightarrow$ We later consider an extension with categorical variables based on NielsenIQ beer types.

Outside option is its own nest; classification runs on inside goods only

Empirical Model

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

$$\delta_{jm} = \beta_p \log \text{price}_{jm} + \beta_s \log \text{size}_j + \beta_n \log \text{units}_j + \varphi_m + \xi_{jm}$$

where size_j is unit size, units_j is units per UPC, and φ_m is a market fixed effect
→ We later consider an extension with categorical variables based on NielsenIQ beer types.

Outside option is its own nest; classification runs on inside goods only

Instruments:

1. Prices: Gandhi-Houde instruments

$$\text{gh}^1_{jm} = \frac{1}{J_m} \sum_{j' \in \mathcal{J}_m} (\log \text{size}_j - \log \text{size}_{j'})^2 \quad \text{and} \quad \text{gh}^2_{jm} = \frac{1}{J_m} \sum_{j' \in \mathcal{J}_m} (\log \text{units}_j - \log \text{units}_{j'})^2$$

→ How much the other products sold in market m differ from product j

Empirical Model

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

$$\delta_{jm} = \beta_p \log \text{price}_{jm} + \beta_s \log \text{size}_j + \beta_n \log \text{units}_j + \varphi_m + \xi_{jm}$$

where size_j is unit size, units_j is units per UPC, and φ_m is a market fixed effect
→ We later consider an extension with categorical variables based on NielsenIQ beer types.

Outside option is its own nest; classification runs on inside goods only

Instruments:

1. Prices: Gandhi-Houde instruments

$$\text{gh}^1_{jm} = \frac{1}{J_m} \sum_{j' \in \mathcal{J}_m} (\log \text{size}_j - \log \text{size}_{j'})^2 \quad \text{and} \quad \text{gh}^2_{jm} = \frac{1}{J_m} \sum_{j' \in \mathcal{J}_m} (\log \text{units}_j - \log \text{units}_{j'})^2$$

→ How much the other products sold in market m differ from product j

Outline

1. Empirical model

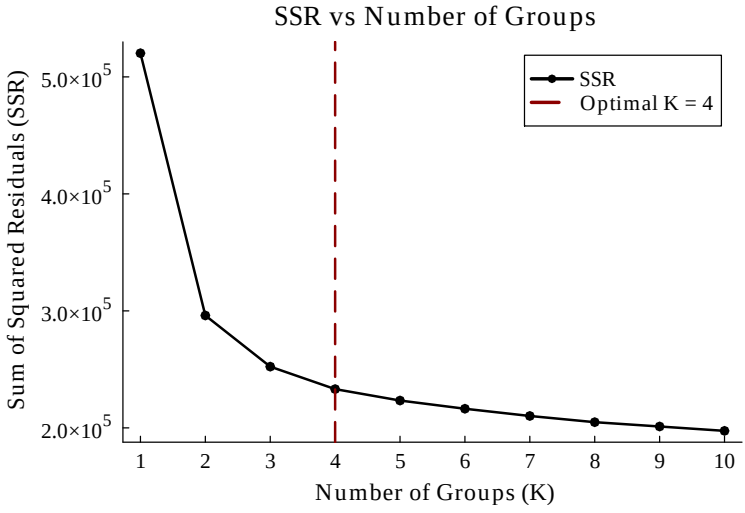
2. Practical aspects and extensions

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 Out-of-sample group validation
   - 4.4 Second-step estimation
   - 4.5 Comparison with alternative empirical models

First-step: Choosing the number of groups

Elbow method suggests $K^* = 4$



First-step: Group characteristics

Characteristics	Unconditional values	Conditional values by group			
		1	2	3	4
Unit size (ml)	354.882	473.176	354.882	354.882	354.882
Units per UPC	6.000	6.000	6.000	6.000	6.000
Unit price	10.209	9.857	10.156	12.367	10.509
Price per ml	0.004	0.003	0.004	0.004	0.005
Share of domestic beer	0.794	0.820	0.773	0.830	0.861
Share of ale	0.158	0.225	0.098	0.291	0.377
Share of regular beer	0.442	0.469	0.430	0.442	0.326
Share of stout and porter	0.015	0.019	0.011	0.018	0.046
Share of light beer	0.369	0.266	0.446	0.238	0.239
Share of other types	0.016	0.022	0.015	0.010	0.012
Share of top ten brands	0.624	0.460	0.755	0.382	0.334
# of products	2806	659	242	1102	803
Label		Regular Lager	Light Beer	Special /Craft	Boutique Craft

Notes: Rows 1–4 report medians, rows 5–11 report means.

First-step: Group characteristics

Characteristics	Unconditional values	Conditional values by group			
		1	2	3	4
Unit size (ml)	354.882	473.176	354.882	354.882	354.882
Units per UPC	6.000	6.000	6.000	6.000	6.000
Unit price	10.209	9.857	10.156	12.367	10.509
Price per ml	0.004	**0.003**	0.004	0.004	0.005
Share of domestic beer	0.794	0.820	0.773	0.830	0.861
Share of ale	0.158	**0.225**	0.098	0.291	0.377
Share of regular beer	0.442	**0.469**	0.430	0.442	0.326
Share of stout and porter	0.015	0.019	0.011	0.018	0.046
Share of light beer	0.369	0.266	0.446	0.238	0.239
Share of other types	0.016	0.022	0.015	0.010	0.012
Share of top ten brands	0.624	**0.460**	0.755	0.382	0.334
# of products	2806	659	242	1102	803
Label		Regular Lager	Light Beer	Special /Craft	Boutique Craft

Notes: Rows 1–4 report medians, rows 5–11 report means.

First-step: Group characteristics

Characteristics	Unconditional values	Conditional values by group			
		1	2	3	4
Unit size (ml)	354.882	473.176	354.882	354.882	354.882
Units per UPC	6.000	6.000	6.000	6.000	6.000
Unit price	10.209	9.857	10.156	12.367	10.509
Price per ml	0.004	0.003	**0.004**	0.004	0.005
Share of domestic beer	0.794	0.820	0.773	0.830	0.861
Share of ale	0.158	0.225	0.098	0.291	0.377
Share of regular beer	0.442	0.469	0.430	0.442	0.326
Share of stout and porter	0.015	0.019	0.011	0.018	0.046
Share of light beer	0.369	0.266	**0.446**	0.238	0.239
Share of other types	0.016	0.022	0.015	0.010	0.012
Share of top ten brands	0.624	0.460	**0.755**	0.382	0.334
# of products	2806	659	242	1102	803
Label		Regular Lager	Light Beer	Special /Craft	Boutique Craft

Notes: Rows 1–4 report medians, rows 5–11 report means.

First-step: Group characteristics

Characteristics	Unconditional values	Conditional values by group			
		1	2	3	4
Unit size (ml)	354.882	473.176	354.882	354.882	354.882
Units per UPC	6.000	6.000	6.000	6.000	6.000
Unit price	10.209	9.857	10.156	12.367	10.509
Price per ml	0.004	0.003	0.004	0.004	0.005
Share of domestic beer	0.794	0.820	0.773	0.830	0.861
Share of ale	0.158	0.225	0.098	**0.291**	**0.377**
Share of regular beer	0.442	0.469	0.430	0.442	0.326
Share of stout and porter	0.015	0.019	0.011	0.018	0.046
Share of light beer	0.369	0.266	0.446	**0.238**	**0.239**
Share of other types	0.016	0.022	0.015	0.010	0.012
Share of top ten brands	0.624	0.460	0.755	**0.382**	**0.334**
# of products	2806	659	242	1102	803
Label		Regular Lager	Light Beer	Special /Craft	Boutique Craft

Notes: Rows 1–4 report medians, rows 5–11 report means.

First-step: Group characteristics

Characteristics	Unconditional values	Conditional values by group			
		1	2	3	4
Unit size (ml)	354.882	473.176	354.882	354.882	354.882
Units per UPC	6.000	6.000	6.000	6.000	6.000
Unit price	10.209	9.857	10.156	**12.367**	**10.509**
Price per ml	0.004	0.003	0.004	**0.004**	**0.005**
Share of domestic beer	0.794	0.820	0.773	0.830	0.861
Share of ale	0.158	0.225	0.098	**0.291**	**0.377**
Share of regular beer	0.442	0.469	0.430	0.442	0.326
Share of stout and porter	0.015	0.019	0.011	0.018	**0.046**
Share of light beer	0.369	0.266	0.446	**0.238**	**0.239**
Share of other types	0.016	0.022	0.015	0.010	0.012
Share of top ten brands	0.624	0.460	0.755	**0.382**	**0.334**
# of products	2806	659	242	1102	803
Label		Regular Lager	Light Beer	Special /Craft	Boutique Craft

Notes: Rows 1–4 report medians, rows 5–11 report means.

Brands by group

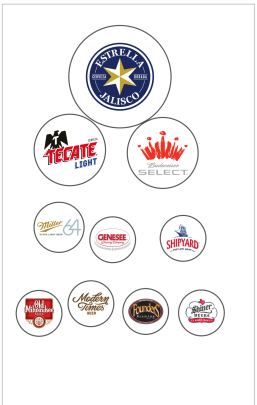
Group 1



Group 2



Group 3



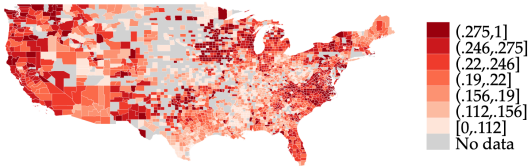
Group 4



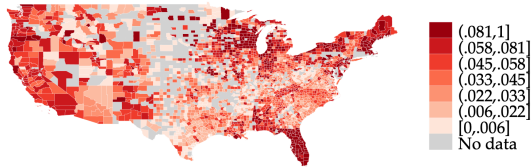
Logos are trademarks of their respective owners

Spatial correlation of county-level market shares by group

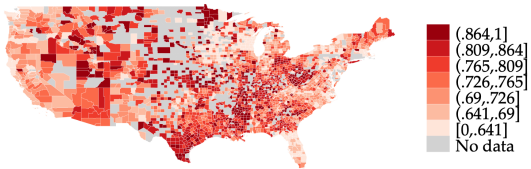
Regular lager



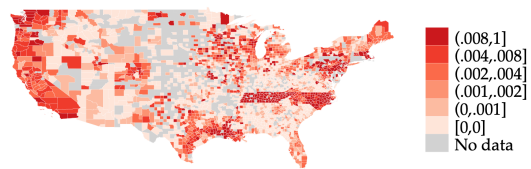
Special / craft



Light beer

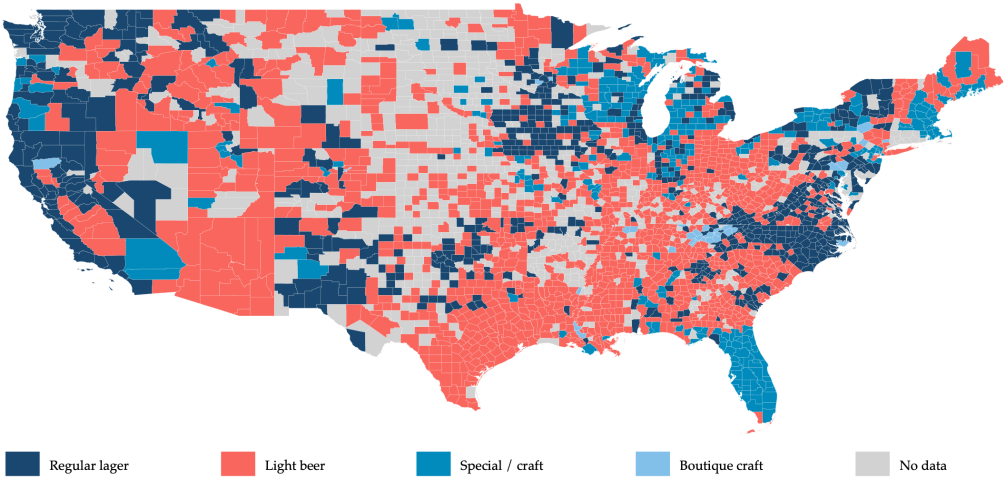


Boutique craft

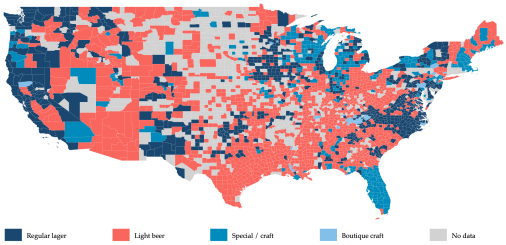


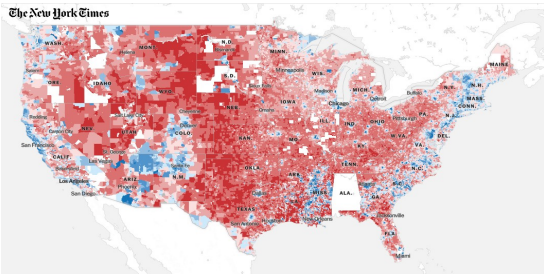
Most popular group (relative to national trends)

High degree of spatial correlation



Group popularity correlated with political outcomes!





Correlation between Light Beer share and 2024 Republican vote share: 0.39

Main takeaway: capturing unobserved drivers of demand

Relationship between alcohol preferences and political ideology has been established in the past

→ e.g., Yakovlev and Guessford (2013)



Alcohol Consumption and Political Ideology: What's Party Got to Do with It?*

Published online by Cambridge University Press: **10 October 2013**

Pavel A. Yakovlev and Walter P. Guessford Text

Show author details

Main takeaway: capturing unobserved drivers of demand

Relationship between alcohol preferences and political ideology has been established in the past
→ e.g., Yakovlev and Guessford (2013)



Alcohol Consumption and Political Ideology: What's Party Got to Do with It?*

Published online by Cambridge University Press: **10 October 2013**

Pavel A. Yakovlev and Walter P. Guessford Text

Show author details

Main takeaway: The Data-driven NL can capture unobserved drivers of demand

- As long as they can be described (or well approximated) by a grouping structure
  - In our case, political ideology interacted with nest tastes:

$$\iota_{km} = \mathbb{1}\{m = \text{republican state}\} \times \text{republican's taste for nest } k$$

    → e.g. Democrats like IPAs, Republicans like light beer
  - ι_{km} is a nest-by-market shifter, absorbed by λ_{km}

Outline

1. Empirical model

2. Practical aspects and extensions

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 Out-of-sample group validation
   - 4.4 Second-step estimation
   - 4.5 Comparison with alternative empirical models

Group validation: Bud Light boycott

Testing out-of-sample nest substitution

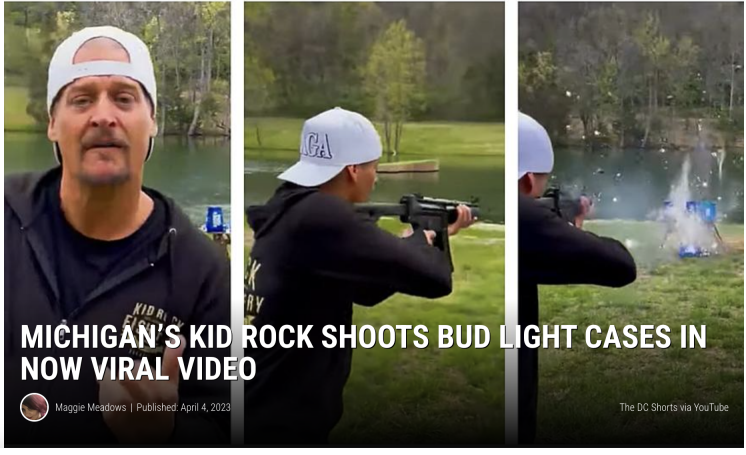
In April 2023, Dylan Mulvaney—an actress, social media influencer, and transgender woman—released a social media promotion of Bud Light



Group validation: Bud Light boycott

Testing out-of-sample nest substitution

The promotion sparked a backlash and calls for a boycott



Group validation: Bud Light boycott

Testing out-of-sample nest substitution

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





Group validation: Bud Light boycott

Testing out-of-sample nest substitution

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

Group validation: Bud Light boycott

Testing out-of-sample nest substitution

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

Hypothesis: if our nests capture substitution patterns

$\Longrightarrow$ Boycott should shift demand toward "Light Beer" products more than toward other nests

Group validation: Bud Light boycott

Testing out-of-sample nest substitution

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

Hypothesis: if our nests capture substitution patterns

$\Longrightarrow$ Boycott should shift demand toward "Light Beer" products more than toward other nests

Test: Empirical model

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

$\Longrightarrow$ augments our state-product panel with 2023 data

Group validation: Bud Light boycott

Testing out-of-sample nest substitution

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

Hypothesis: if our nests capture substitution patterns

$\Longrightarrow$ Boycott should shift demand toward "Light Beer" products more than toward other nests

Test: Empirical model

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

$\Longrightarrow$ augments our state-product panel with 2023 data

Because of UPC-market fixed effects κ_{jm}

$\Longrightarrow$ exploit within-product temporal variation not used in the classification of groups

Group validation: DiD results

Dependent Variable: Log Share								
	First stage		DiD				Placebo	
t=2023	-0.358***	(0.005)	-0.402***	(0.005)	-0.444***	(0.008)	-0.355***	(0.010)
Log Prices	-0.234***	(0.054)	-0.244***	(0.054)	-0.259***	(0.054)	-0.217***	(0.054)
Treatment (Budlight)	**-0.155****	(0.026)						
Treatment same group			0.425***	(0.008)				
Regular beer × t=2023					0.177***	(0.012)		
Light beer × t=2023					0.468***	(0.010)		
Non-lager/craft × t=2023					0.002	(0.013)		
Boutique craft × t=2023					0.017	(0.020)		
Placebo Regular beer × t=2023							-0.020	(0.014)
Placebo Light beer × t=2023							-0.003	(0.014)
Placebo Non-lager/craft × t=2023							-0.004	(0.014)
Placebo Boutique craft × t=2023							0.009	(0.015)
Observations	177784		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES	
Exclude Bud Light	No		YES		YES		YES	
Baseline group	**Non-Budlight**		Other groups		Outside option		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Group validation: DiD results

Dependent Variable: Log Share								
	First stage		DiD				Placebo	
t=2023	-0.358***	(0.005)	-0.402***	(0.005)	-0.444***	(0.008)	-0.355***	(0.010)
Log Prices	-0.234***	(0.054)	-0.244***	(0.054)	-0.259***	(0.054)	-0.217***	(0.054)
Treatment (Budlight)	-0.155***	(0.026)						
Treatment same group			**0.425***	(0.008)				
Regular beer × t=2023					0.177***	(0.012)		
Light beer × t=2023					0.468***	(0.010)		
Non-lager/craft × t=2023					0.002	(0.013)		
Boutique craft × t=2023					0.017	(0.020)		
Placebo Regular beer × t=2023							-0.020	(0.014)
Placebo Light beer × t=2023							-0.003	(0.014)
Placebo Non-lager/craft × t=2023							-0.004	(0.014)
Placebo Boutique craft × t=2023							0.009	(0.015)
Observations	177784		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES	
Exclude Bud Light	No		YES		YES		YES	
Baseline group	Non-Budlight		**Other groups**		Outside option		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Group validation: DiD results

Dependent Variable: Log Share

	First stage		DiD				Placebo	
t=2023	-0.358***	(0.005)	-0.402***	(0.005)	-0.444***	(0.008)	-0.355***	(0.010)
Log Prices	-0.234***	(0.054)	-0.244***	(0.054)	-0.259***	(0.054)	-0.217***	(0.054)
Treatment (Budlight)	-0.155***	(0.026)						
Treatment same group			0.425***	(0.008)				
Regular beer × t=2023					0.177***	(0.012)		
Light beer × t=2023					**0.468*** **	(0.010)		
Non-lager/craft × t=2023					0.002	(0.013)		
Boutique craft × t=2023					0.017	(0.020)		
Placebo Regular beer × t=2023							-0.020	(0.014)
Placebo Light beer × t=2023							-0.003	(0.014)
Placebo Non-lager/craft × t=2023							-0.004	(0.014)
Placebo Boutique craft × t=2023							0.009	(0.015)
Observations	177784		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES	
Exclude Bud Light	No		YES		YES		YES	
Baseline group	Non-Budlight		Other groups		**Outside option**		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Outline

1. Empirical model

2. Practical aspects and extensions

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 Out-of-sample group validation
   - 4.4 Second-step estimation
   - 4.5 Comparison with alternative empirical models

Second-step results: Structural parameters and matrix of elasticities

Rich heterogeneity in substitution patterns

	OLS	IV
log unit size	0.070 (0.006)***	0.339 (0.025)***
log upc quantity	0.080 (0.005)***	0.441 (0.026)***
log price	-0.054 (0.006)***	-0.416 (0.033)***
$\sigma_{\text{regular/lager}}$	0.142 (0.001)***	0.523 (0.014)***
$\sigma_{\text{light beer}}$	0.343 (0.002)***	0.807 (0.016)***
$\sigma_{\text{special/craft}}$	-0.015 (0.001)***	0.336 (0.013)***
$\sigma_{\text{boutique craft}}$	-0.320 (0.001)***	0.042 (0.015)***
Mean Own-price Elasticity	1.351 (0.003)	-3.561 (0.013)
Mean Cross-price Elasticity	-0.003 (0.008)	0.011 (0.033)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	

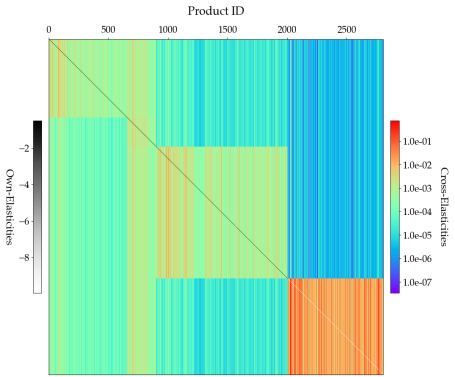
Literature range of average beer own-price elasticities is $[-6.1, -3.6]$

(Hausman et al., 1994; Pinkse and Slade, 2004; Slade, 2004; Hellerstein, 2008; Miller and Weinberg, 2017)

Second-step results: Structural parameters and matrix of elasticities

Rich heterogeneity in substitution patterns

	OLS	IV
log unit size	0.070 (0.006)***	0.339 (0.025)***
log upc quantity	0.080 (0.005)***	0.441 (0.026)***
log price	-0.054 (0.006)***	-0.416 (0.033)***
$\sigma_{\text{regular/lager}}$	0.142 (0.001)***	0.523 (0.014)***
$\sigma_{\text{light beer}}$	0.343 (0.002)***	0.807 (0.016)***
$\sigma_{\text{special/craft}}$	-0.015 (0.001)***	0.336 (0.013)***
$\sigma_{\text{boutique craft}}$	-0.320 (0.001)***	0.042 (0.015)***
Mean Own-price Elasticity	1.351 (0.003)	-3.561 (0.013)
Mean Cross-price Elasticity	-0.003 (0.008)	0.011 (0.033)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	



Literature range of average beer own-price elasticities is $[-6.1, -3.6]$

(Hausman et al., 1994; Pinkse and Slade, 2004; Slade, 2004; Hellerstein, 2008; Miller and Weinberg, 2017)

Outline

1. Empirical model

2. Practical aspects and extensions

3. Consistency, statistical properties, and Monte Carlo

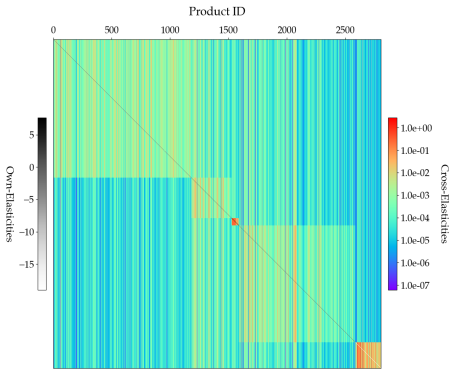
4. Empirical application: Demand for beer
   - 4.1 Data description and empirical model
   - 4.2 First-step estimation: Classification
   - 4.3 Out-of-sample group validation
   - 4.4 Second-step estimation
   - 4.5 Comparison with alternative empirical models

Naive Nested Logit: Using NielsenIQ "Beer Type" as Nests

	OLS	IV
log unit size	0.054 (0.005)***	0.686 (0.041)***
log upc quantity	0.059 (0.004)***	0.694 (0.042)***
log price	-0.039 (0.005)***	-0.753 (0.050)***
σ_{ale}	-0.029 (0.001)***	0.239 (0.012)***
$\sigma_{\text{light beer}}$	0.087 (0.001)***	0.348 (0.010)***
σ_{other}	-0.356 (0.002)***	-0.098 (0.015)***
$\sigma_{\text{regular/lager}}$	0.080 (0.001)***	0.322 (0.010)***
$\sigma_{\text{stout/porter}}$	-0.296 (0.001)***	0.040 (0.017)**
Mean Own-price Elasticity	0.360 (0.003)	-3.752 (0.104)
Mean Cross-price Elasticity	-0.001 (0.003)	0.007 (0.111)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	

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

	OLS	IV
log unit size	0.054 (0.005)***	0.686 (0.041)***
log upc quantity	0.059 (0.004)***	0.694 (0.042)***
log price	-0.039 (0.005)***	-0.753 (0.050)***
σ_{ale}	-0.029 (0.001)***	0.239 (0.012)***
$\sigma_{\text{light beer}}$	0.087 (0.001)***	0.348 (0.010)***
σ_{other}	-0.356 (0.002)***	-0.098 (0.015)***
$\sigma_{\text{regular/lager}}$	0.080 (0.001)***	0.322 (0.010)***
$\sigma_{\text{stout/porter}}$	-0.296 (0.001)***	0.040 (0.017)**
Mean Own-price Elasticity	0.360 (0.003)	-3.752 (0.104)
Mean Cross-price Elasticity	-0.001 (0.003)	0.007 (0.111)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	



Bud Light boycott: Comparing Data-driven vs. Naive NL

Dependent Variable: Log Share										
	Data-driven NL				Naive NL				Interaction	
t=2023	-0.402***	(0.005)	-0.444***	(0.008)	-0.322***	(0.006)	-0.444***	(0.008)	-0.374***	(0.007)
Log Prices	-0.244***	(0.054)	-0.259***	(0.054)	-0.241***	(0.054)	-0.266***	(0.054)	-0.254***	(0.054)
Treatment same group	**0.425***	(0.008)								
Regular beer × t=2023			0.177***	(0.012)						
Light beer × t=2023			0.468***	(0.010)						
Non-lager/craft × t=2023			0.002	(0.013)						
Boutique craft × t=2023			0.017	(0.020)						
Naive light beer × t=2023					**0.131***	(0.016)	0.254***	(0.017)		
Naive ale × t=2023							0.029**	(0.012)		
Naive other × t=2023							0.013	(0.043)		
Naive lager × t=2023							0.254***	(0.012)		
Naive stout × t=2023							-0.040*	(0.021)		
DD Light only									0.387***	(0.009)
Naive Light only									0.033	(0.024)
DD and Naive Light									0.424***	(0.012)
Observations	172778		172778		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES		YES	
Baseline group	Other groups		Outside option		Outside option		Outside option		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Bud Light boycott: Comparing Data-driven vs. Naive NL

Dependent Variable: Log Share										
	Data-driven NL				Naive NL				Interaction	
t=2023	-0.402***	(0.005)	-0.444***	(0.008)	-0.322***	(0.006)	-0.444***	(0.008)	-0.374***	(0.007)
Log Prices	-0.244***	(0.054)	-0.259***	(0.054)	-0.241***	(0.054)	-0.266***	(0.054)	-0.254***	(0.054)
Treatment same group	0.425***	(0.008)								
Regular beer × t=2023			0.177***	(0.012)						
Light beer × t=2023			**0.468*** **	(0.010)						
Non-lager/craft × t=2023			0.002	(0.013)						
Boutique craft × t=2023			0.017	(0.020)						
Naive light beer × t=2023					0.131***	(0.016)	**0.254*** **	(0.017)		
Naive ale × t=2023							0.029**	(0.012)		
Naive other × t=2023							0.013	(0.043)		
Naive lager × t=2023							0.254***	(0.012)		
Naive stout × t=2023							-0.040*	(0.021)		
DD Light only									0.387***	(0.009)
Naive Light only									0.033	(0.024)
DD and Naive Light									0.424***	(0.012)
Observations	172778		172778		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES		YES	
Baseline group	Other groups		Outside option		Outside option		Outside option		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Bud Light boycott: Comparing Data-driven vs. Naive NL

Dependent Variable: Log Share

	Data-driven NL				Naive NL				Interaction	
t=2023	-0.402***	(0.005)	-0.444***	(0.008)	-0.322***	(0.006)	-0.444***	(0.008)	-0.374***	(0.007)
Log Prices	-0.244***	(0.054)	-0.259***	(0.054)	-0.241***	(0.054)	-0.266***	(0.054)	-0.254***	(0.054)
Treatment same group	0.425***	(0.008)								
Regular beer × t=2023			0.177***	(0.012)						
Light beer × t=2023			0.468***	(0.010)						
Non-lager/craft × t=2023			0.002	(0.013)						
Boutique craft × t=2023			0.017	(0.020)						
Naive light beer × t=2023					0.131***	(0.016)	0.254***	(0.017)		
Naive ale × t=2023							0.029**	(0.012)		
Naive other × t=2023							0.013	(0.043)		
Naive lager × t=2023							0.254***	(0.012)		
Naive stout × t=2023							-0.040*	(0.021)		
DD Light only									**0.387*****	(0.009)
Naive Light only									**0.033**	(0.024)
DD and Naive Light									**0.424*****	(0.012)
Observations	172778		172778		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES		YES	
Baseline group	Other groups		Outside option		Outside option		Outside option		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** 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

$$\delta_{jm} = \beta_p \text{log price}_{jm} + \beta_s \text{log size}_j + \beta_n \text{log units}_j + \varphi_m + \xi_{jm},$$

and μ_{ijm} is a random coefficient on log price,

$$\mu_{ijm} = \left(\sigma_p \nu_i\right) \log \text{p}_{jm}, \quad \nu_i \sim N(0, 1),$$

where σ_p is the standard deviation of the random price coefficient

Mixed Logit Structural estimates

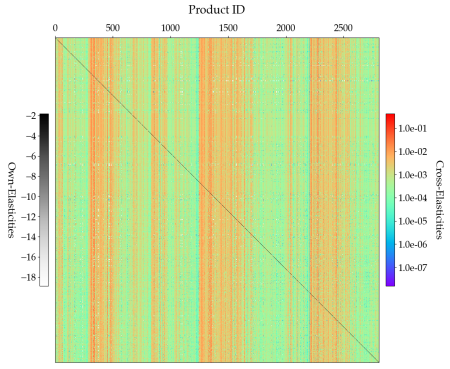
For this setting and specification, Mixed Logit looks close to a simple Logit

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 Coefficient	σ_p
log price	0.002*** (0.0003)
Own-price elasticity	-4.683 (2.863)
Cross-price elasticity	0.002 (0.005)
Market Fixed Effects	✓
IV Type	Gandhi-Houde
Number of Products	2806
Number of Observations	64497

Mixed Logit Structural estimates

For this setting and specification, Mixed Logit looks close to a simple Logit

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 Coefficient	σ_p
log price	0.002*** (0.0003)
Own-price elasticity	-4.683 (2.863)
Cross-price elasticity	0.002 (0.005)
Market Fixed Effects	✓
IV Type	Gandhi-Houde
Number of Products	2806
Number of Observations	64497



Mixed Logit

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

Dependent Variable: Log Share								
	Continuous		DiD					
t=2023	-0.289***	(0.008)	-0.290***	(0.008)	-0.312***	(0.006)	-0.449***	(0.008)
Log Prices	-0.163*	(0.084)	-0.240***	(0.054)				
$\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023	**-7.048****	(1.585)						
Above median $\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023			**-0.031****	(0.011)				
Top quartile $\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023					**-0.003**	(0.012)		
Quartile 1 $\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023							**0.219****	(0.014)
Quartile 2 $\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023							0.085***	(0.013)
Quartile 3 $\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023							0.107***	(0.014)
Quartile 4 $\mathcal{E}^i_{\text{Bud Light}} \times$ t=2023							**0.134****	(0.013)
Observations	107404		172778		172778		172778	
UPC-Market FE	YES		YES		YES		YES	
Exclude Bud Light	YES		YES		YES		YES	
Baseline group	Outside option		Below median		Other quartiles		Outside option	

Notes: Heteroskedasticity-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Conclusions

Proposed a two-step estimator to estimate nesting structure

Showed conditions for consistency of nests and estimator

Empirical application based on US beer market shows:

- Reasonable results and rich substitution patterns with **minimal data and computational requirements**
- Nests correlated with geography/political preferences
  → Data-driven Nested Logit can capture unobservable factors of demand
- For this application, our method outperforms Naive Nested Logit and Mixed Logit in one 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, identical to the blue bus except for its 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, identical to the blue bus except for its 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, identical to the blue bus except for its 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

$$\begin{aligned}\log \mathbb{P}_{jm} &= \frac{\beta_x x_{jm} + \beta_p p_{jm} + \xi_{jm}}{\sigma_{k(j)}} + \lambda_{k(j)m} \\ &= \underbrace{\beta_{k(j),x}}_{\frac{\beta_x}{\sigma_k}} x_{jm} + \underbrace{\beta_{k(j),p}}_{\frac{\beta_p}{\sigma_k}} p_{jm} + \lambda_{k(j)m} + \xi_{jm},\end{aligned}$$

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 $\lambda_{k(j)m}$ and $\lambda_{k(j')m}$

One-step nest estimation

With exogenous prices, combine steps 1 + 2 by solving the constrained problem

$$\underset{\substack{k(1),\ldots,k(J)\\ \beta_1,\ldots,\beta_K,\lambda_{11},\ldots,\lambda_{KM}\\ \beta,\sigma_1,\ldots,\sigma_K}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\left(\log\left(\frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}\right)-(x_{jm}\beta_k+\lambda_{km})\right)^2$$

$$\text{s.t.}\quad \beta_k=\frac{\beta}{\sigma_k}\qquad\text{and}\qquad \lambda_{km}=\frac{\sigma_k-1}{\sigma_k}\log\frac{\mathbb{P}_{km}}{\mathbb{P}_{0m}}$$

One-step nest estimation

With exogenous prices, combine steps 1 + 2 by solving the constrained problem

$$\underset{\substack{k(1),\ldots,k(J)\\ \beta_1,\ldots,\beta_K,\lambda_{11},\ldots,\lambda_{KM}\\ \beta,\sigma_1,\ldots,\sigma_K}}{\arg\min} \sum_{m=1}^{M}\sum_{j=1}^{J}\left(\log\left(\frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}\right)-(x_{jm}\beta_k+\lambda_{km})\right)^2$$

$$\text{s.t.}\quad \beta_k=\frac{\beta}{\sigma_k}\qquad\text{and}\qquad \lambda_{km}=\frac{\sigma_k-1}{\sigma_k}\log\frac{\mathbb{P}_{km}}{\mathbb{P}_{0m}}$$

- **Not robust:** with a nest-market unobservable η_{km}, the one-step estimator is inconsistent
  → The two-step absorbs η_{km} in λ_{km}
  → The NL formula for λ_{km} also fails under higher-level nests and nested CES
- **No theory:** λ_{km} is non-linear in σ_k and depends on the partition
  → GFE consistency is only established for linear models with free group-market effects

Nash-Bertrand pricing

Assume marginal costs can be written as

$$\log mc_{jm} = \gamma_x^{c(j)} x_{jm} + \gamma_z^{c(j)} z_{jm} + \alpha_{c(j)m} + \omega_{jm}$$

Nash-Bertrand pricing

Assume marginal costs can be written as

$$\log mc_{jm} = \gamma_x^{c(j)} x_{jm} + \gamma_z^{c(j)} z_{jm} + \alpha_{c(j)m} + \omega_{jm}$$

Under Bertrand–Nash competition with single-product firms, with $|\mathcal{E}^j_{jm}|$ the own-price elasticity:

$$p_{jm} = \frac{1}{1 - \frac{1}{|\mathcal{E}^j_{jm}|}} mc_{jm}$$

Nash-Bertrand pricing

Assume marginal costs can be written as

$$\log mc_{jm} = \gamma_x^{c(j)} x_{jm} + \gamma_z^{c(j)} z_{jm} + \alpha_{c(j)m} + \omega_{jm}$$

Under Bertrand–Nash competition with single-product firms, with $|\mathcal{E}_{jm}^{j}|$ the own-price elasticity:

$$p_{jm} = \frac{1}{1 - \frac{1}{|\mathcal{E}_{jm}^{j}|}} mc_{jm}$$

With nested logit demand and log prices,

$$|\mathcal{E}_{jm}^{j}| = \frac{|\beta_p|}{\sigma_{k(j)}}\Big(1 - \sigma_{k(j)}\mathbb{P}_{jm} - (1 - \sigma_{k(j)})\mathbb{P}_{j|k(j),m}\Big) = \frac{|\beta_p|}{\sigma_{k(j)}} + O_p(J^{-1})$$

Nash-Bertrand pricing

Assume marginal costs can be written as

$$\log mc_{jm} = \gamma_x^{c(j)} x_{jm} + \gamma_z^{c(j)} z_{jm} + \alpha_{c(j)m} + \omega_{jm}$$

Under Bertrand–Nash competition with single-product firms, with $|\mathcal{E}^j_{jm}|$ the own-price elasticity:

$$p_{jm} = \frac{1}{1 - \frac{1}{|\mathcal{E}^j_{jm}|}} mc_{jm}$$

With nested logit demand and log prices,

$$|\mathcal{E}^j_{jm}| = \frac{|\beta_p|}{\sigma_{k(j)}}\Big(1 - \sigma_{k(j)}\mathbb{P}_{jm} - (1 - \sigma_{k(j)})\mathbb{P}_{j|k(j),m}\Big) = \frac{|\beta_p|}{\sigma_{k(j)}} + O_p(J^{-1})$$

Substituting into prices:

$$\log p_{jm} = \gamma_x^{c(j)} x_{jm} + \gamma_z^{c(j)} z_{jm} + \underbrace{\left(\alpha_{c(j)m} - \log\left(1 - \frac{\sigma_{k(j)}}{|\beta_p|}\right)\right)}_{\kappa_{g(j)m}} + \omega_{jm} + O_p(J^{-1})$$

Classification algorithm

1. Let $\boldsymbol{\theta}^0 \equiv (\beta_1^0, \dots, \beta_K^0, \lambda_{11}^0, \dots, \lambda_{KM}^0)$ be a starting value.

Classification algorithm

1. Let $\boldsymbol{\theta}^0 \equiv (\beta_1^0, \dots, \beta_K^0, \lambda_{11}^0, \dots, \lambda_{KM}^0)$ be a starting value.
2. For $\boldsymbol{\theta}^s$, compute for all $j = 1, \dots, J$:

$$k(j)^{s+1} = \underset{k \in \{1,\dots,K\}}{\arg\min} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_{k,p}^s \hat{p}_{jm} + \beta_{k,x}^s x_{jm} + \lambda_{km}^s) \right)^2,$$

to obtain nesting structure $\mathcal{B}^{s+1}$.

Classification algorithm

1. Let $\boldsymbol{\theta}^0 \equiv (\beta_1^0, \dots, \beta_K^0, \lambda_{11}^0, \dots, \lambda_{KM}^0)$ be a starting value.
2. For $\boldsymbol{\theta}^s$, compute for all $j = 1, \dots, J$:

$$k(j)^{s+1} = \underset{k \in \{1,\dots,K\}}{\arg\min} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_{k,p}^s \hat{p}_{jm} + \beta_{k,x}^s x_{jm} + \lambda_{km}^s) \right)^2,$$

   to obtain nesting structure $\mathcal{B}^{s+1}$.

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

$$\boldsymbol{\theta}^{s+1} = \underset{\beta_1,\dots,\beta_K,\lambda_{11},\dots,\lambda_{KM}}{\arg\min} \sum_{j=1}^{J} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_{k(j),p}^{s+1} \hat{p}_{jm} + \beta_{k(j),x}^{s+1} x_{jm} + \lambda_{k(j)m}^{s+1}) \right)^2$$

Classification algorithm

1. Let $\boldsymbol{\theta}^0 \equiv (\beta_1^0, \dots, \beta_K^0, \lambda_{11}^0, \dots, \lambda_{KM}^0)$ be a starting value.
2. For $\boldsymbol{\theta}^s$, compute for all $j = 1, \dots, J$:

$$k(j)^{s+1} = \underset{k \in \{1,\dots,K\}}{\arg\min} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_{k,p}^s \hat{p}_{jm} + \beta_{k,x}^s x_{jm} + \lambda_{km}^s) \right)^2,$$

   to obtain nesting structure $\mathcal{B}^{s+1}$.

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

$$\boldsymbol{\theta}^{s+1} = \underset{\beta_1,\dots,\beta_K,\lambda_{11},\dots,\lambda_{KM}}{\arg\min} \sum_{j=1}^{J} \sum_{m=1}^{M} \left(\log \mathbb{P}_{jm} - (\beta_{k(j),p}^{s+1} \hat{p}_{jm} + \beta_{k(j),x}^{s+1} x_{jm} + \lambda_{k(j)m}^{s+1}) \right)^2$$

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

GMM Cannot Jointly Recover Groups and Parameters

The GMM estimator solves:

$$\min_{\theta,\, k(\cdot)} \; \bar{m}(\theta, k(\cdot))^\top W \, \bar{m}(\theta, k(\cdot)), \qquad \bar{m}(\theta, k(\cdot)) = \frac{1}{JM} \sum_{j=1}^{J} \sum_{m=1}^{M} \begin{pmatrix} \xi_{jm}\left(\theta_{k(j)}\right) z_{jm} \\ \xi_{jm}\left(\theta_{k(j)}\right) x_{jm} \end{pmatrix}.$$

In the just-identified case, for *any* candidate $k(\cdot)$, one can *always* find $\hat{\theta}(k(\cdot))$ such that:

$$\bar{m}\left(\hat{\theta}(k(\cdot)),\ k(\cdot)\right) = 0$$

$\Longrightarrow$ the objective is zero for every $k(\cdot)$

$\Longrightarrow$ $k(\cdot)$ and parameters cannot be jointly identified via GMM

GMM Cannot Jointly Recover Groups and Parameters

The GMM estimator solves:

$$\min_{\theta,\, k(\cdot)} \; \bar{m}(\theta, k(\cdot))^\top W \, \bar{m}(\theta, k(\cdot)), \qquad \bar{m}(\theta, k(\cdot)) = \frac{1}{JM} \sum_{j=1}^{J} \sum_{m=1}^{M} \begin{pmatrix} \xi_{jm}\left(\theta_{k(j)}\right) z_{jm} \\ \xi_{jm}\left(\theta_{k(j)}\right) x_{jm} \end{pmatrix}.$$

In the just-identified case, for *any* candidate $k(\cdot)$, one can *always* find $\hat{\theta}(k(\cdot))$ such that:

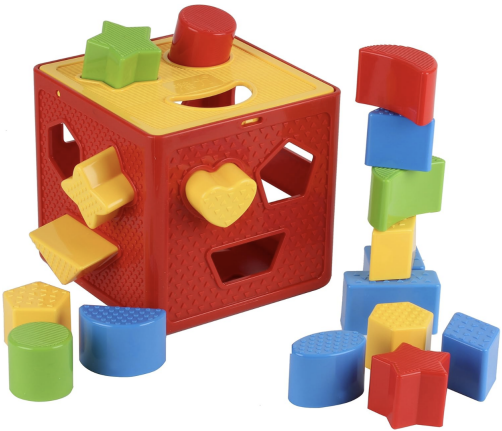
$$\bar{m}\Big(\hat{\theta}(k(\cdot)),\; k(\cdot)\Big) = 0$$

$\Longrightarrow$ the objective is zero for every $k(\cdot)$
$\Longrightarrow$ $k(\cdot)$ and parameters cannot be jointly identified via GMM

Result extends to over-identified case: for any grouping $k(\cdot)$, there exist parameters θ such that the GMM objective converges to zero in probability. Groups remain unidentified.

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

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

$$y_{jm} = \alpha^*_{k(j)} + \xi_{jm}, \quad k(j) \in \{1, 2\},$$

  where α^* denotes true values and α candidate values.

- We characterize the misclassification probability:

$$\mathbb{P}\left(\widehat{k}(j) = 2 \middle| k(j) = 1\right) = \mathbb{P}\left(\left(\overline{y}_j - \alpha_2\right)^2 < \left(\overline{y}_j - \alpha_1\right)^2 \middle| k(j) = 1\right).$$

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

$$\mathbb{P}\left(\overline{\xi}_j > \frac{\alpha_1 + \alpha_2}{2} - \alpha_1^*\right) \quad = \quad 1 - \Phi\left(\frac{\sqrt{M}}{\sigma}\left(\frac{\alpha_1 + \alpha_2}{2} - \alpha_1^*\right)\right) \to 0 \text{ as } M \to \infty,$$

- **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!

Robustness: $K = 3$

Brands by nest

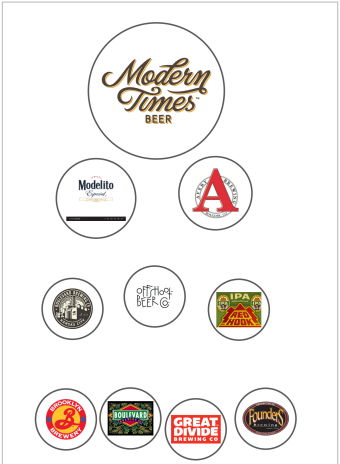
Group 1



Group 2



Group 3



Back

Robustness: $K = 3$

Second-step results

	OLS	IV
log unit size	0.059 (0.006)***	0.681 (0.040)***
log upc quantity	0.067 (0.004)***	0.824 (0.042)***
log price	-0.067 (0.005)***	-1.038 (0.056)***
σ_1	0.272 (0.001)***	0.968 (0.024)***
σ_2	0.047 (0.001)***	0.649 (0.022)***
σ_3	-0.204 (0.001)***	0.391 (0.022)***
Mean Own-price Elasticity	-0.457 (0.001)	-2.010 (0.002)
Mean Cross-price Elasticity	0.001 (0.003)	0.002 (0.006)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	

Robustness: $K = 5$

Brands by nest

Group 1


Miller
HIGH LIFE
PACIFICO
CLARA
XX
BUSCH
KEYSTONE
LIGHT
TECATE
Sol
ICEHOUSE
DOS EQUIS
XX
LAGER
CERVEZA
ESPECIAL

Robustness: $K = 5$

Second-step results: one σ at the boundary, one not significant

	OLS	IV
log unit size	0.087 (0.007)***	0.196 (0.027)***
log upc quantity	0.103 (0.005)***	0.241 (0.027)***
log price	-0.037 (0.007)***	-0.084 (0.036)**
σ_1	0.202 (0.002)***	0.698 (0.016)***
σ_2	0.407 (0.002)***	1.000 (0.019)***
σ_3	0.060 (0.001)***	0.516 (0.015)***
σ_4	-0.096 (0.001)***	0.340 (0.015)***
σ_5	-0.442 (0.002)***	0.029 (0.018)
Mean Own-price Elasticity	-0.057 (0.001)	-0.729 (0.004)
Mean Cross-price Elasticity	0.000 (0.002)	0.003 (0.011)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	

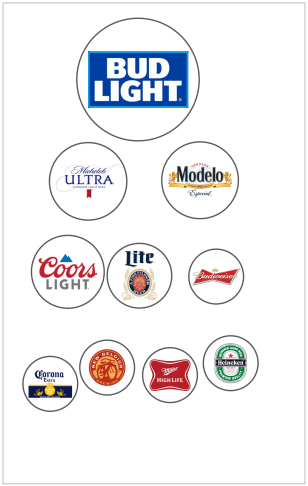
Robustness: classification with “Beer Type” dummies

Brands by nest, $K^* = 3$

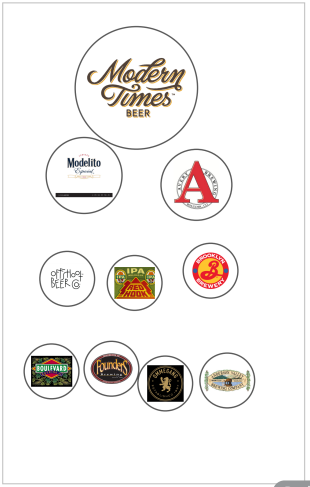
Group 1



Group 2



Group 3



Logos are trademarks of their respective owners

Robustness: classification with "Beer Type" dummies

Second-step results, $K^* = 3$: own-price elasticity -2.3, low end of the literature range

	OLS	IV
log unit size	0.037 (0.006)***	0.088 (0.024)***
log upc quantity	0.040 (0.005)***	0.122 (0.024)***
log price	-0.040 (0.006)***	-0.146 (0.031)***
σ_1	0.046 (0.001)***	0.303 (0.009)***
σ_2	0.269 (0.001)***	0.578 (0.010)***
σ_3	-0.209 (0.001)***	0.031 (0.009)***
Mean Own-price Elasticity	-0.276 (0.001)	-2.343 (0.006)
Mean Cross-price Elasticity	0.000 (0.002)	0.005 (0.015)
Market Fixed Effects	✓	✓
IV Type		GH_{units}, GH_{size}
Number of Products	2806	
Number of Observations	64497	