Data-Driven Nests in Discrete Choice Models

Milena Almagro (Minneapolis Fed and Chicago Booth) and Elena Manresa (NYU)

May 4, 2021

Motivation

Discrete Choice Models (I)

- Models of discrete choice are the workhorse in demand estimation with random utility
- Based on the random utility framework
    - Utility is driven by observable characteristics and an idiosyncratic taste shock
    - Agents choose the alternative with highest utility
- If idiosyncratic shocks are $\sim$ Type I EV $\implies$ Multinomial logit:
    - Closed form solutions of choice probability
    - Low number of parameters
    - Generates unrealistic substitution patterns

Discrete Choice Models (II)

- A number of different models have been proposed to alleviate these “undesirable features”
- Random Coefficients (RC): Logit with heterogeneity in preferences across consumers
  - Flexible substitution patterns.
  - Computationally expensive: non-linear optimization, no closed-form demand.
  - Distributional assumptions on heterogeneity.
- Nested Logit (NL): Natural extension of Multinomial Logit
  - Closed form solutions, parsimony (linear IV regression), interpretability.
  - Less flexible substitution patterns.
  - Nests need to be specified ex-ante.

This Paper: Estimate Groups in Nested Logit

- Methodology to estimate the nest structure as well as preference parameters
- No assumption on the nest structure $\Longrightarrow$ recovered from the data
- Two-step estimation procedure:
    1. Use k-means to estimate the nest structure
    2. Estimate the utility parameters as if the groups where known
- Identification of the nests and statistical properties.
- Monte Carlo Simulations
- Empirical Application: US Automobile Sale Data.

Related Literature

- **Discrete Choice Models:** McFadden (1978), Cardell (1997), Kovach & Tserenjigmid (2020)
- **Empirical Models with Nesting Structures:** Goldberg (1995), Einav (2007), Grennan (2013), Ciliberto & Williams (2014), Conlon & Rao (2016), Miller & Weinberg (2017)
- **Group Fixed Effect Estimator:** Han & Moon (2010), Bonhomme & Manresa (2015), Phillips et al (...)
- **Alternative Grouping Structure:** Fosgerau, Monardo, & De Palma (2021), Hortacsu, Lieber, Monardo & de Paula (ongoing)

Consumer choice model

Consumer choice model: a recap

- Consider the following indirect utility model:

$$V_{ij} = \delta_j + \varepsilon_{ij}$$

  for agent i when choosing j.

- When $\epsilon_{ij} \sim$ Type I EV, choice probabilities are given by:

$$\mathbb{P}_j = \frac{\exp(\delta_j)}{\sum_{j'} \exp(\delta_{j'})}$$

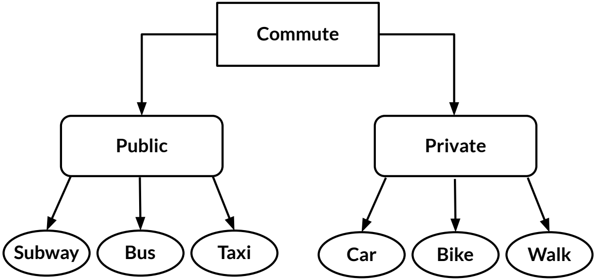
- With K groups and $(\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$ have cumulative distribution $\sim \exp\left(-\sum_{k=1}^{K} \left(\sum_{j \in B_K} e^{-\frac{\epsilon_j}{\sigma^{k(j)}}}\right)^{\sigma^{k(j)}}\right)$:

$$\mathbb{P}_j = \frac{e^{\frac{\delta_j}{\sigma^{k(j)}}} \left(\sum_{d \in B_k} e^{\frac{\delta_d}{\sigma^{k(d)}}}\right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_{l(j)}} e^{\frac{\delta_d}{\sigma^{l(d)}}}\right)^{\sigma^{l}}}$$

Nested Logit as Sequential Choice

Choose nest, then alternative within nest:

$$\mathbb{P}_j = \frac{e^{\frac{\delta_j}{\sigma^{k(j)}}}}{\sum_{d \in B_k} e^{\frac{\delta_d}{\sigma^{k(d)}}}} \frac{\left(\sum_{d \in B_k} e^{\frac{\delta_d}{\sigma^{k(d)}}} \right)^{\sigma^{k(j)}}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_d}{\sigma^{l(d)}}} \right)^{\sigma^l}} = \mathbb{P}_{j|k(j)} \mathbb{P}_{k(j)}$$



Nested Logit: Substitution Patterns

Elasticities:

- For Logit:

$$E_j^q = -\mathbb{P}_q \frac{\partial \delta_q}{\partial p_q} p_q$$

- For Nested Logit:

$$E_j^q = \begin{cases} -\mathbb{P}_q \frac{\partial \delta_q}{\partial p_q} p_q & \text{if } q \in B_{k'(q)} \neq B_{k(j)} \\ (\sigma^{k(j)} - 1)\mathbb{P}_{q|k(j)} \frac{\partial \delta_q}{\partial p_q} p_q - \mathbb{P}_q \frac{\partial \delta_q}{\partial p_q} p_q & \text{if } q \in B_{k(j)} \end{cases}$$

$\Longrightarrow$ Elasticity Multinomial Logit $\leq$ Elasticity Nested Logit (within nest).
$\Longrightarrow$ Products within same nest closer substitutes than across nests.
$\Longrightarrow$ More substitution as σ^k decreases.
$\Longrightarrow$ $(1 - \sigma^{k(j)}) \in [0, 1]$ can be interpreted as correlation within nest.

Blue Bus - Red Bus

Identification

Identification of Groups (I)

Let

$$\text{IV}^k \equiv \sum_{d \in B_k} e^{\frac{\delta_d}{\sigma^{k(d)}}} \quad \text{and} \quad \text{IV} \equiv \sum_{l=1}^{K} \Big(\sum_{d \in B_l} e^{\frac{\delta_d}{\sigma^{k(d)}}} \Big)^{\sigma^l}$$

then

$$\mathbb{P}_j = \frac{e^{\frac{\delta_j}{\sigma^{k(j)}}} \left(\sum_{d \in B_k} e^{\frac{\delta_d}{\sigma^{k(d)}}} \right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\delta_d}{\sigma^{l(d)}}} \right)^{\sigma^l}} = \frac{e^{\frac{\delta_j}{\sigma^{k(j)}}} \left(IV^{k(j)} \right)^{\sigma^{k(j)}-1}}{IV}$$

Taking logs:

$$\log \mathbb{P}_j = \frac{\delta_j}{\sigma^{k(j)}} + (\sigma^{k(j)} - 1) \log \text{IV}^{k(j)} - \log \text{IV}.$$

We assume $\delta_0 = 0$ and $k(0) = \{0\}$ as in Berry (1994). It follows:

$$\log \frac{\mathbb{P}_j}{\mathbb{P}_0} = \frac{\delta_j}{\sigma^{k(j)}} + (\sigma^{k(j)} - 1) \log \text{IV}^{k(j)}$$

Identification of Groups (II)

We now assume linear utility in one observable component:

$$\delta_j = \beta x_j$$

Substituting inside choice probabilities

$$\log \frac{\mathbb{P}_j}{\mathbb{P}_0} = \frac{\beta}{\sigma^{k(j)}} x_j + (\sigma^{k(j)} - 1) \log IV^{k(j)}$$

Denote

$$\beta^{k(j)} = \frac{\beta}{\sigma^{k(j)}} \qquad \text{and} \qquad \lambda^{k(j)} = (\sigma^{k(j)} - 1) \log IV^{k(j)}$$

We obtain the following equation:

$$\log \frac{\mathbb{P}_j}{\mathbb{P}_0} = \beta^{k(j)} x_j + \lambda^{k(j)}$$

$\implies \beta^k$ and λ^k common to all products within the same nest!

Identification of Groups (III)

$$\log \frac{\mathbb{P}_j}{\mathbb{P}_0} = \beta^{k(j)} x_j + \lambda^{k(j)}$$

Intuition: Assume products j and j' have same x's

1. If $\log \frac{\mathbb{P}_j}{\mathbb{P}_0}$ is equal to $\log \frac{\mathbb{P}'_j}{\mathbb{P}_0} \implies j$ and j' are in the same group.
2. If $\log \frac{\mathbb{P}_j}{\mathbb{P}_0}$ is different from $\log \frac{\mathbb{P}'_j}{\mathbb{P}_0} \implies j$ and j' are **not** in the same group.

Identification of Groups (III)

$$\log \frac{\mathbb{P}_j}{\mathbb{P}_0} = \beta^{k(j)} x_j + \lambda^{k(j)}$$

Intuition: Assume products j and j' have same x's

1. If $\log \frac{\mathbb{P}_j}{\mathbb{P}_0}$ is equal to $\log \frac{\mathbb{P}_j'}{\mathbb{P}_0} \implies j$ and j' are in the same group.
2. If $\log \frac{\mathbb{P}_j}{\mathbb{P}_0}$ is different from $\log \frac{\mathbb{P}_j'}{\mathbb{P}_0} \implies j$ and j' are **not** in the same group.

Remarks:

1. We think of $\beta^{k(j)}$ and $\lambda^{k(j)}$ as group-specific slope and intercept in a regression equation.
2. Identification of the groups is hence obtained without fully imposing the structure of the model.

The first step not only recovers groups but also $\{\beta^1, \ldots, \beta^K, \lambda^1, \ldots, \lambda^K\}$.

Recall

$$\log \frac{\mathbb{P}_j}{\mathbb{P}_0} = \frac{\delta_j}{\sigma^{k(j)}} + \lambda^{k(j)}, \qquad \lambda^k = (\sigma^k - 1) \log IV_k = (\sigma^k - 1) \log \sum_{j \in B_k} e^{\frac{\delta_j}{\sigma^k}} \quad \text{and} \quad \beta^k = \frac{\beta}{\sigma^k}$$

Then, σ^k and β are jointly identified from the following equations:

$$\lambda^k = \frac{\sigma^k - 1}{\sigma^k} \log \Big(\sum_{j \in B_k} \frac{\mathbb{P}_j}{\mathbb{P}_0} \Big) \quad \text{and} \quad \beta^k = \frac{\beta}{\sigma^k}$$

Estimation

Relevant Empirical Models

In what follows, we assume that covariates x are exogenous but allow for endogeneity of prices p.

We consider two different models of indirect utility:

1. A panel data framework with product fixed effects:

$$\delta_{jm} = \beta_p p_{jm} + x_{jm}\beta_x + \xi_j + \nu_{jm},$$

   where

$$\mathbb{E}[\nu_{jm}|p_{j1}, \ldots, p_{jM}, x_{j1}, \ldots, x_{jM}, \xi_j] = 0$$

2. Panel data with exogenous shifters:

$$\delta_{jm} = \beta_p p_{jm} + x_{jm}\beta_x + \nu_{jm},$$

   where there exists z_{jm} such that

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

Outline

1 Motivation

2 Consumer choice model

3 Identification

4 Estimation
- Case 1: Panel Data
- Case 2: Exogenous Shifters

5 Statistical Properties

6 Choosing the Number of Groups

7 Monte Carlo

8 Application: US Automobile Data

Panel Data: Estimation

Recall, indirect utility is defined as:

$$\delta_{jm} = \beta_p p_{jm} + x_{jm}\beta_x + \xi_j + \nu_{jm},$$

where

$$\mathbb{E}[\nu_{jm} | p_{j1}, \ldots, p_{jM}, x_{j1}, \ldots, x_{jM}, \xi_j] = 0$$

Therefore,

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \frac{\beta_p p_{jm} + x_{jm}\beta_x + \xi_j + \nu_{jm}}{\sigma^{k(j)}} + \lambda^{k(j),m} = \beta_p^{k(j)} p_{jm} + x_{jm}\beta_x^{k(j)} + \tilde{\xi}_j + \tilde{\nu}_{jm} + \lambda_m^{k(j)},$$

where $\tilde{\xi}_j = \frac{\xi_j}{\sigma^{k(j)}}$ and $\tilde{\nu}_{jm} = \frac{\nu_{jm}}{\sigma^{k(j)}}$

We demean data to remove fixed effects $\tilde{\xi}_j$

$$\overline{\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}} = \beta_p^{k(j)} \bar{p}_{jm} + \bar{x}_{jm}\beta_x^{k(j)} + \bar{\lambda}_m^{k(j)} + \bar{\nu}_{jm},$$

where $\bar{y}$ indicates demeaned variables.

First Step: Classification Algorithm

We propose the following classification algorithm based on Bonhomme and Manresa (2015):

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

First Step: Classification Algorithm

We propose the following classification algorithm based on Bonhomme and Manresa (2015):

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

$$k(j)^{s+1} = \underset{k \in \{1,\ldots,K\}}{\arg\min} \sum_{m=1}^{M} \left(\overline{\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}} - \left(\bar{x}_{jm} \beta^{k,s} + \lambda_m^{k,s} \right) \right)^2,$$

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

First Step: Classification Algorithm

We propose the following classification algorithm based on Bonhomme and Manresa (2015):

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

$$k(j)^{s+1} = \underset{k\in\{1,\ldots,K\}}{\arg\min} \sum_{m=1}^{M} \left(\log\frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} - \left(\bar{x}_{jm}\beta^{k,s} + \lambda_m^{k,s}\right)\right)^2,$$

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

3. Compute:

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

First Step: Classification Algorithm

We propose the following classification algorithm based on Bonhomme and Manresa (2015):

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

$$k(j)^{s+1} = \arg\min_{k \in \{1,\ldots,K\}} \sum_{m=1}^{M} \left(\overline{\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}} - \left(\bar{x}_{jm} \beta^{k,s} + \lambda_m^{k,s} \right) \right)^2,$$

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

3. Compute:

$$(\beta^{1,s+1}, \ldots, \beta^{K,s+1}, \lambda_1^{K,s+1}, \ldots, \lambda_M^{K,s+1}) =$$
$$\arg\min_{\beta^1,\ldots,\beta^K,\lambda_1^1,\ldots,\lambda_M^K} \sum_{j=1}^{J} \sum_{m=1}^{M} \left(\overline{\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}}} - \left(\bar{x}_{jm} \beta^{k(j),s+1} + \lambda_m^{k(j),s+1} \right) \right)^2$$

4. Repeat until convergence of parameters.

Second Step: Linear Regression

Based on the estimated classification from the first step, we follow Berry (1994).

Under normalization $\delta_0 = 0$ and $k(0) = \{0\}$

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_0} = \delta_{jm} + (1 - \sigma^{k(j)}) \log \mathbb{P}_{jm|k(j)} + \nu_{jm}$$

Substituting the expression for δ_j:

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

Linear regression equation on x_{jm} and $\log \mathbb{P}_{jm|k(j)}$:

- Construct $\mathbb{P}_{j|k(j)}$ based on estimated groups from first step $\{\hat{B}_k\}_{k=1}^{K}$:

$$\hat{\mathbb{P}}_{jm|k} = \frac{\mathbb{P}_{jm}}{\sum_{j \in \hat{B}_k} \mathbb{P}_j}$$

- Simultaneity problem: $\implies$ Instrument $\hat{\mathbb{P}}_{jm|k(j)}$ using second order moments of exogenous x_{jm}.

Outline

1. Motivation
2. Consumer choice model
3. Identification
4. Estimation
   - Case 1: Panel Data
   - Case 2: Exogenous Shifters
5. Statistical Properties
6. Choosing the Number of Groups
7. Monte Carlo
8. Application: US Automobile Data

Control Function Approach: Estimation

Assume indirect utility model is described as:

$$\delta_{jm} = \beta_p p_{jm} + x_{jm}\beta_x + \nu_{jm}$$

If $\mathbb{E}[\nu_{jm} p_{jm}] \neq 0$, the algorithm outlined before does not consistently recover the groups.

To overcome this issue, we use a Control Function Approach defined in Petrin and Train (2010).

We require the existence of z_{jm} such that

$$\mathbb{E}[\nu_{jm} | x_{jm}, z_{jm}] = 0.$$

Control Function Approach

Concretely, there is an unobservable confounder μ_{jm} such that:

$$p_{jm} = f(x_{jm}, z_{jm}, \mu_{jm}) \quad \text{and} \quad \nu_{jm} = g(\mu_{jm}, \varepsilon_{jm}),$$

for which we assume that

$$p_{jm} \perp\!\!\!\perp \nu_{jm} | \mu_{jm}.$$

For simplicity we also assume:

$$p_{jm} = f(x_{jm}, z_{jm}; \gamma) + \mu_{jm} \quad \text{and} \quad \nu_{jm} = CF(\mu_{jm}) + \varepsilon_{jm}$$

Include $CF(\mu_{jm})$ as part of indirect utility :

$$\begin{aligned} \delta_{jm} &= \beta_p p_{jm} + x_{jj}\beta_x + \nu_{jm} \\ &= \beta_p p_{jm} + x_{jm}\beta_x + CF(\mu_{jm}) + \varepsilon_{jm}, \end{aligned}$$

with $\mathbb{E}[\varepsilon_{jm} | p_{jm}, x_{jm}, \mu_{jm}] = 0$.

Control Function Approach (cont)

Substituting δ_{jm} inside choice probabilities:

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \beta_p^{k(j)} p_{jm} + x_j \beta_x^{k(j)} + \widetilde{CF}(\mu_{jm}) + \tilde{\lambda}_m^{k(j)} + \tilde{\varepsilon}_{jm},$$

which is a known expression with p_{jm}, x_{jm} and μ_{jm} as observable covariates.

This expression motivates the following steps:

1. Project p_{jm} on exogenous variables (z_{jm}, x_{jm}) to estimate μ_{jm}

$$\hat{\mu}_{jm} = p_{jm} - \hat{f}(x_{jm}, z_{jm})$$

2. Include $\hat{\mu}_{jm}$ in our classification algorithm as a control for the confounder between ν_{jm} and p_{jm}.
3. Follow step 2 as if the groups are known.

Statistical Properties

Statistical Properties: First Step

- Let us consider the following simplified model:

$$\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m} = \beta_p^{k(j)} p_{jm} + x_{jm}\beta_x^{k(j)} + \lambda_m^{k(j)} + \nu_{jm}$$

  with $\mathbb{E}[\nu_{jm}|p_{j1}, \ldots, p_{jM}, x_{j1}, \ldots, x_{jM}, \lambda_1^1, \ldots \lambda_M^K] = 0$.

- Build upon results in Bonhomme and Manresa (2015).
- Work in progress: allow for product fixed effects and projection of prices.

Three key assumptions

- **Group separation**. For simplicity, assume simplest model:

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \lambda^{k(j)} + \nu_{jm}, \qquad \text{with } k \in \{1,2\},\ \lambda^2 > \lambda^1,\ \nu_{jm} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

It follows

$$\mathbb{P}(\hat{k}(j) = 2 | k(j) = 1) = \mathbb{P}\Big(\sum_{m=1}^{M} (\lambda^1 + \nu_{jm} - \lambda^2)^2 < \sum_{m=1}^{M} (\lambda^1 + \nu_{jm} - \lambda^1)^2 \Big) = \mathbb{P}(\bar{\nu}_j > \lambda^2 - \lambda^1) = 1 - \Phi\Big(\sqrt{M}\big(\frac{\lambda^2 - \lambda^1}{2}\big)\Big)$$

Three key assumptions

- **Group separation**. For simplicity, assume simplest model:

$$\log \frac{\mathbb{P}_{jm}}{\mathbb{P}_{0m}} = \lambda^{k(j)} + \nu_{jm}, \qquad \text{with } k \in \{1,2\},\ \lambda^2 > \lambda^1,\ \nu_{jm} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

  It follows

$$\mathbb{P}(\hat{k}(j) = 2 | k(j) = 1) = \mathbb{P}\Big(\sum_{m=1}^{M} (\lambda^1 + \nu_{jm} - \lambda^2)^2 < \sum_{m=1}^{M} (\lambda^1 + \nu_{jm} - \lambda^1)^2 \Big) = \mathbb{P}(\bar{\nu}_j > \lambda^2 - \lambda^1) = 1 - \Phi\Big(\sqrt{M}\big(\frac{\lambda^2 - \lambda^1}{2}\big)\Big)$$

- **Rank condition**: Variation in x at the intersection of any group with true groups
- **Exponential tails** and limited market dependence on the error term

Super consistency

- It can be shown:

$$\mathbb{P}\Big(\sup_{j\in\{1,2,\ldots,J\}} |\widehat{k}(j) - k(j)| > 0\Big) = o(1) + o(JM^{-\delta})$$

  for any $\delta > 0$, as J and M go to infinity.

- Both J and M grow to infinity, but M can grow at a much lower rate!
- "Super consistency" of group estimation $\implies$ standard inference in the second step.

Choosing the Number of Groups

Choosing K: Cross-Validation with Elbow Method

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

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

For all $k \in \mathcal{K}$:

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

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

- Take average across N folds:

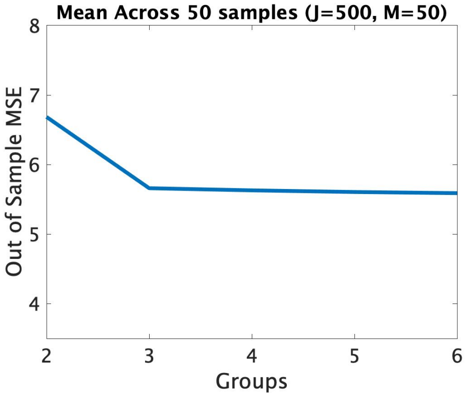
$$MSE(k) = \frac{1}{N} \sum_{n=1}^{N} MSE_n(k)$$

- Choose k according to

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

Cross Validation: Simulation Results

groups = 3, # folds = 5, # MC samples = 50



Moving M

Monte Carlo

Monte Carlo Design (I)

- Indirect utility δ_{jm} is given by

$$\delta_{jm} = \beta_p p_{jm} + \beta_1 x_{jm,1} + \beta_2 x_{jm,2} + \xi_j + \nu_{jm},$$

where $p_{jm,1}$ are prices and $(x_{jm,1}, x_{jm,2})$ are exogenous covariates. We set:

- $p_{jm,1} = \tilde{p}_{jm} + \xi_{j,p}$, with:
  - $\tilde{p}_{jm,1} \overset{i.i.d.}{\sim} \mathcal{N}\Big(k(j) \cdot \arctan(m+1), 1\Big)$
  - $\begin{bmatrix} \xi_{j,p} \\ \xi_j \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\Big)$
- $x_{jm,1}, x_{jm,2} \overset{i.i.d.}{\sim} \mathcal{N}\Big(k(j) \cdot (-1)^{k(j)} \cdot \arctan(m+1), 1\Big)$
- $\mathbb{E}[\nu_{jm} | p_{j1}, x_{j1,1}, x_{j1,2}, \ldots, p_{jM}, x_{jM,1}, x_{jM,2}, \xi_j] = 0$ with $\nu_{jm} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$
- $\beta_p = -1$ and $\beta_1 = \beta_2 = 1$

- Number of groups $K = 3$ with $\sigma_1 = 0.3, \sigma_2 = 0.5, \sigma_3 = 0.7$.

Monte Carlo Design: Outcome variables

We leverage the closed form solution of Nested Logit models.

Construct IV_m^k as follows:

$$IV_{k,m} = \Big(\sum_{d \in B_k} e^{\frac{\delta_{dm}}{\sigma^k}} \Big)$$

Finally, log probabilities are given by:

$$\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m} = \frac{1}{\sigma^{k(j)}} \delta_{jm} + (\sigma^{k(j)} - 1) \log IV_m^{k(j)}$$

Results: K = 3, I = 1000, B = 500

M	J	% Matches	Time (s)	True	β_p -1	β_1 1	β_2 1	σ_1 0.3	σ_2 0.5	σ_3 0.7
10	100	0.911	2	Mean β	-0.898	0.896	0.894	0.254	0.425	0.627
				Std β	0.122	0.131	0.131	0.065	0.063	0.063
50	100	1.000	13	Mean β	-0.956	0.957	0.958	0.286	0.476	0.669
				Std β	0.058	0.059	0.059	0.030	0.031	0.030
100	100	1.000	66	Mean β	-0.971	0.970	0.971	0.291	0.485	0.679
				Std β	0.046	0.046	0.046	0.024	0.024	0.024
10	500	0.879	14	Mean β	-0.912	0.904	0.903	0.264	0.429	0.629
				Std β	0.078	0.080	0.080	0.038	0.037	0.039
50	500	0.996	273	Mean β	-0.959	0.959	0.958	0.287	0.478	0.671
				Std β	0.047	0.047	0.047	0.024	0.024	0.024
100	500	1.000	710	Mean β	-0.967	0.967	0.967	0.290	0.483	0.677
				Std β	0.044	0.044	0.044	0.023	0.023	0.022
10	1000	0.870	25	Mean β	-0.903	0.898	0.897	0.267	0.427	0.625
				Std β	0.054	0.056	0.056	0.026	0.027	0.026
50	1000	0.988	471	Mean β	-0.963	0.963	0.963	0.289	0.478	0.673
				Mean std	0.0381	0.0382	0.0381	0.0190	0.0192	0.0192
100	1000	1.000	2145	Mean β	-0.976	0.976	0.975	0.292	0.487	0.683
				Std β	0.047	0.047	0.047	0.014	0.024	0.033

Application: US Automobile Data

US Automobile Data

We use US Automobile data from BLP (1995).[1]

Information on (essentially) all models marketed between 1971 and 1990.

Models both enter and exit over this period $\Longrightarrow$ unbalanced panel.

Total sample size is 2217 model/years representing 557 distinct models.

We set different years as different markets.

[1] We use data from the `R`-package `hdm` developed by Chernozhukov, Hansen & Spindler (2019)

Product Characteristics

Description of product characteristics:

- *log share*: log of market shares
- *price*: deflated price to 1983 dollars using CPI
- *mpd*: miles per dollar
- *air*: air conditioning
- *mpg*: miles per gallon rating
- *space*:size (measured as length times width)
- *hpwt*: the ratio of horsepower to weight (in HP per 10 lbs)

Panel Construction

- We consider an unbalanced panel of cars with:
  - At least five years of data.
  - At least three consecutive years.
- We are left with 82 products.
- We adapt our classification algorithm to allow for “missing data”:
  $\Longrightarrow$ Products can enter and exit over time.
  $\Longrightarrow$ Group of products can also enter and exit over time!

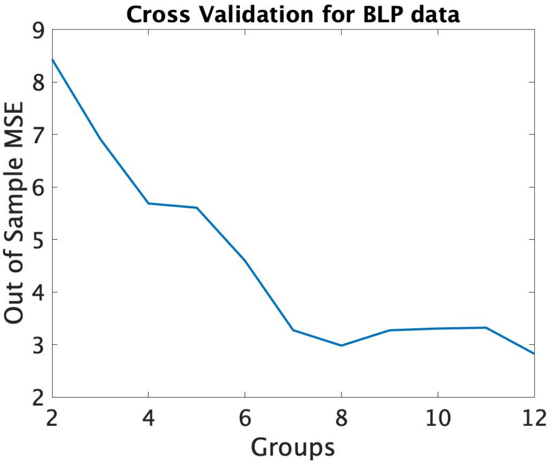
Statistics

Statistics of subsample of cars (N=82)

	Mean	Std. Dev.	Median	Min	Max	t-stat
Price	-.147	7.911	-2.532	-6.601	43.351	-1.06
Miles per Dollar	2.349	.513	2.376	1.352	3.805	2.78
AC	.299	.409	0	0	1	0.49
Miles per Gallon	2.214	.46	2.195	1.38	3.42	1.45
Space	1.266	.187	1.223	.951	1.711	0.13
Horse Power	.407	.069	.386	.308	.727	-0.23
Market Share	.001	.001	.001	0	.004	0.00
Yearly Observations	9.085	4.264	7	5	20	10.42
Year Entry	1980	5.261	1983	1971	1986	-4.62
Year Exit	1989	.88	1990	1988	1990	20.41

Full sample

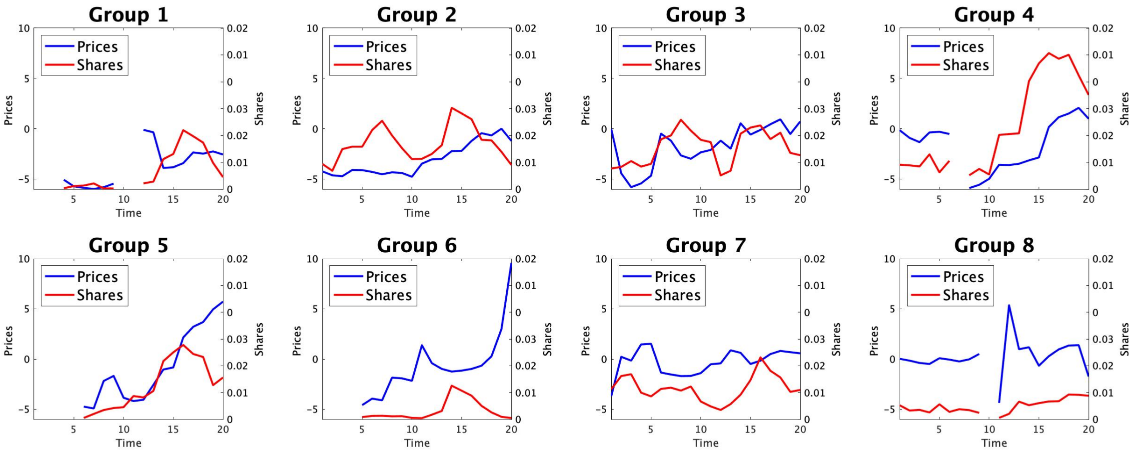
BLP Application: Choosing the number of groups



BLP Application: First-step Group Characteristics

	Mean	Std.	1	2	3	4	5	6	7	8
Shares	0.001	0.001	0.004	0.009	0.008	0.012	0.006	0.002	0.006	0.002
Price	-0.741	6.898	-3.679	-3.077	-1.694	-1.621	-0.688	-0.610	-0.292	0.211
Log HP	-0.940	0.183	-1.054	-0.973	-0.984	-0.976	-0.942	-0.876	-0.953	-0.915
Log Miles per $	0.767	0.320	0.919	0.623	0.653	0.650	0.823	0.641	0.610	0.642
AC	0.277	0.448	0.072	0.315	0.259	0.268	0.132	0.144	0.303	0.267
Log Space	0.239	0.164	0.096	0.315	0.259	0.282	0.176	0.180	0.303	0.281
Type of car			Subc.	Compact Mid-size	Mid-size Luxury	Luxury	Mid-size Luxury	Sport	Mid-size Full-size	Full-size Luxury
# Products	82		7	11	11	15	12	8	12	6

BLP Application: Evolution of Shares



BLP Application: Second-step Results

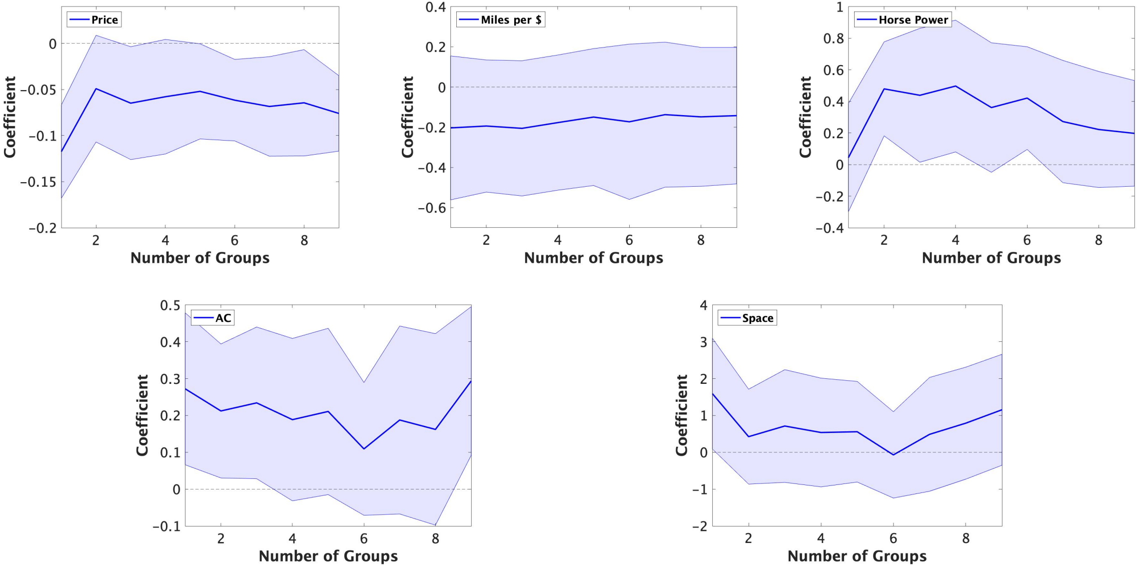
Estimates Preference Parameters

	$\hat{\beta}$	$\sigma_{\hat{\beta}}$
Price	-0.064***	(0.029)
Horse Power	-0.148	(0.176)
Miles per $	0.222	(0.187)
AC	0.1621	(0.133)
Space	0.791	(0.775)

Estimates Within-Nest Correlation

	Group							
	1	2	3	4	5	6	7	8
$\hat{\sigma}$	0.868***	0.596***	0.472***	0.827***	0.722***	0.836***	0.528***	0.572***
$\sigma_{\hat{\sigma}}$	(0.155)	(0.277)	(0.165)	(0.104)	(0.273)	(0.139)	(0.145)	(0.173)
F 1st stage	50.673	2.7697	6.241	6.320	6.963	16.311	11.805	11.748

Moving K: Preference Parameters



Moving K: Within-nest Correlations

# Groups	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9
2	***0.525***	***0.372***							
3	***0.619***	**0.614**	***0.590***						
4	***0.693***	**0.660**	***0.499***	***0.459***					
5	**0.816**	**0.681**	***0.573***	***0.547***	***0.362***				
6	**0.807**	***0.759***	***0.601***	***0.355***	***0.237***	***0.213***			
7	**1.235**	**0.850**	**0.837**	**0.704**	**0.659**	**0.526**	**0.330**		
8	**0.868**	**0.836**	**0.827**	**0.722**	**0.596**	***0.572***	***0.528***	***0.472***	
9	**0.965**	**0.758**	***0.729***	***0.676***	**0.644**	0.535	***0.528***	0.439	*-0.160*

Notes: Bold = different from 0 at the 95%, Italic = different from 1 at the 95%.

Conclusion

- Method that simultaneously estimates nests and preference parameters in nested logit models.

Conclusion

- Method that simultaneously estimates nests and preference parameters in nested logit models.
- Two models to overcome endogeneity of prices:
  1. Panel with product fixed effects that are correlated with prices
  2. Panel with exogenous shifters onto which project prices

Conclusion

- Method that simultaneously estimates nests and preference parameters in nested logit models.
- Two models to overcome endogeneity of prices:
    1. Panel with product fixed effects that are correlated with prices
    2. Panel with exogenous shifters onto which project prices
- Monte Carlo simulations:
    - $\sim$ 90% match rate with only 10 and $\sim$ 100% with 100 markets.
    - Biases in preference parameters decrease as number of market increases.

Conclusion

- Method that simultaneously estimates nests and preference parameters in nested logit models.
- Two models to overcome endogeneity of prices:
  1. Panel with product fixed effects that are correlated with prices
  2. Panel with exogenous shifters onto which project prices
- Monte Carlo simulations:
  - $\sim$ 90% match rate with only 10 and $\sim$ 100% with 100 markets.
  - Biases in preference parameters decrease as number of market increases.
- BLP application:
  - Eight groups with separation in prices, car characteristics, and market trends.
  - Wide range of substitution patterns, from very independent to highly correlated.

IIA is not always realistic: Red-bus-Blue-bus problem

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

Assume indirect utility from the two is the same so

$$\mathbb{P}_c = \mathbb{P}_{bb} = \frac{1}{2} \implies \frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$$

IIA is not always realistic: Red-bus-Blue-bus problem

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

Assume indirect utility from the two is the same so

$$\mathbb{P}_c = \mathbb{P}_{bb} = \frac{1}{2} \implies \frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$$

Now a **red bus** is introduced, exactly equal to **blue bus** (but the color) $\implies \frac{\mathbb{P}_{rb}}{\mathbb{P}_{bb}} = 1$

Given IIA, $\frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$. The only consistent model with both is

$$\mathbb{P}_c = \mathbb{P}_{bb} = \mathbb{P}_{rb} = \frac{1}{3}$$

IIA is not always realistic: Red-bus-Blue-bus problem

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

Assume indirect utility from the two is the same so

$$\mathbb{P}_c = \mathbb{P}_{bb} = \frac{1}{2} \implies \frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$$

Now a **red bus** is introduced, exactly equal to **blue bus** (but the color) $\implies \frac{\mathbb{P}_{rb}}{\mathbb{P}_{bb}} = 1$

Given IIA, $\frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$. The only consistent model with both is

$$\mathbb{P}_c = \mathbb{P}_{bb} = \mathbb{P}_{rb} = \frac{1}{3}$$

Is $\mathbb{P}_c = \mathbb{P}_{bb} = \mathbb{P}_{rb} = \frac{1}{3}$ realistic?

IIA is not always realistic: Red-bus-Blue-bus problem

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

Assume indirect utility from the two is the same so

$$\mathbb{P}_c = \mathbb{P}_{bb} = \frac{1}{2} \implies \frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$$

Now a **red bus** is introduced, exactly equal to **blue bus** (but the color) $\implies \frac{\mathbb{P}_{rb}}{\mathbb{P}_{bb}} = 1$

Given IIA, $\frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$. The only consistent model with both is

$$\mathbb{P}_c = \mathbb{P}_{bb} = \mathbb{P}_{rb} = \frac{1}{3}$$

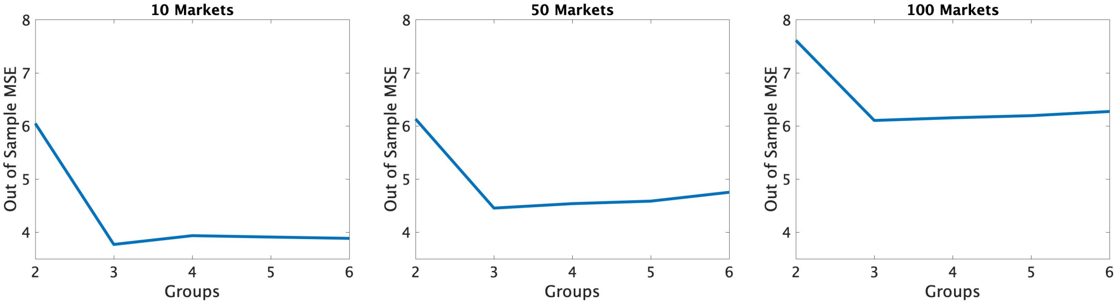
Is $\mathbb{P}_c = \mathbb{P}_{bb} = \mathbb{P}_{rb} = \frac{1}{3}$ realistic? Not really.
If **blue** and **red** only differ in color, we should expect

$$\mathbb{P}_c = \frac{1}{2} \qquad \mathbb{P}_{bb} = \mathbb{P}_{rb} = \frac{1}{4}$$

The ratio $\frac{\mathbb{P}_c}{\mathbb{P}_{bb}}$ should actually change with the introduction of the **red bus**!

Back

groups = 3, # products = 100, # folds = 5



Back

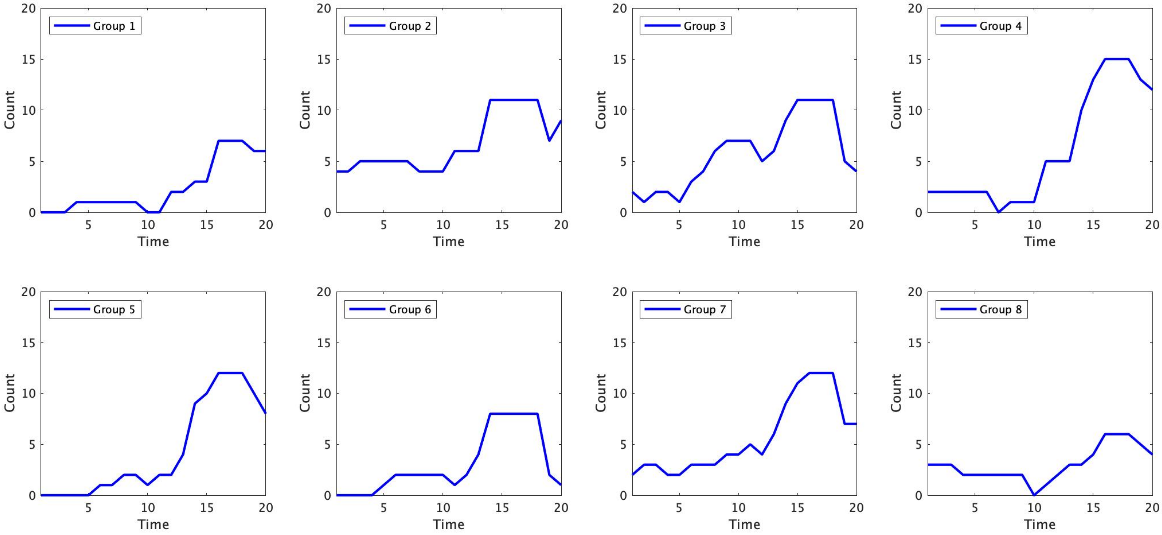
Statistics for Full Sample

Table: Average characteristics of all cars, (N = 557)

	Mean	Std. Dev.	Median	Min	Max	t-stat
Price	.862	8.983	-2.516	-8.368	43.351	1.06
Miles per Dollar	2.175	.641	2.094	1.055	6.437	-2.78
AC	.275	.424	0	0	1	-0.49
Miles per gallon	2.133	.552	2.07	1	5.3	-1.45
Space	1.263	.216	1.223	.79	1.888	-0.13
Horse Power	.409	.098	.385	.207	.888	0.23
Market Share	.001	.001	0	0	0.006	0.00
Yearly Observations	3.899	3.857	2	1	20	-10.42
Entry Year	1980	6.511	1981	1971	1990	4.62
Exit Year	1984	6.101	1986	1971	1990	-20.41

Back

BLP Application: Evolution of Size



BLP Application: First Step Group Fixed Effects