Data-Driven Nests in Discrete Choice Models

Milena Almagro and Elena Manresa

November 27, 2019

Motivation

Motivation

- Models of discrete choice are the workhorse in the IO literature
- Different models of utility maximization
  $\rightarrow$ trade-off between parsimony and flexibility
- A widely used model is one in which idiosyncratic taste shocks $\sim$ Type I EV $\implies$ multinomial logit
  - $(+)$ Closed form solutions of choice probability
  - $(+)$ Low number of parameters
  - (-) Generates unrealistic substitution patterns (more on this later).
- A number of different models have been proposed to alleviate these "undesirable features".

A Solution: Nested Logit

- Nested Logit is a natural extension of Multinomial Logit with many of its good features such as closed form solutions, parsimony, interpretability.
- Nested logit represents consumers as agents that choose sequentially over product categories
- However, Nested Logit has been criticized as sometimes is hard to argue that the relevant level of categorization ex ante is known.

This Paper

- We propose a method to estimate Nested Logit in which we estimate the nests with the rest of the parameters of the model
- We use the fact that there are lots of individuals and relatively few products.
- The method requires observing a cross-section of market shares. This allows for the possibility to learn nonparametrically how substitution patterns change over time as a result of entry/exit or mergers.
- We use k-means, a classification method from machine learning, to estimate the groups in a first step, and then estimate the rest of the parameters.

The Model

The Model with Independent Errors: Multinomial Logit

- Assume agent i receives utility $V_{ij} = \beta x_j + \epsilon_{ij}$ when choosing j.
- Probability that agent i chooses product j

$$\begin{aligned} \mathbb{P}_j &= \mathbb{P}(V_{ij} > V_{ij'} \; \forall j' \neq j) \\ & \mathbb{P}(\beta x_j + \epsilon_{ij} > \beta x_{j'} + \epsilon_{ij'} \; \forall j' \neq j) \\ &= \mathbb{P}(\beta x_j - \beta x_{j'} > \epsilon_{ij'} - \epsilon_{ij} \; \forall j' \neq j) \end{aligned}$$

- When $\epsilon_{ij} \sim$ Type I Extreme Value

$$\implies \epsilon_{ij'} - \epsilon_{ij} \sim \text{ logit distribution}$$

then choice probability has closed form solution

$$\mathbb{P}_j = \frac{\exp(\beta x_j)}{\sum_{j'} \exp(\beta x_{j'})}$$

Independence of Irrelevant Alternatives

Logit errors imply for any two alternatives j and j'

$$\begin{aligned}\frac{\mathbb{P}_j}{\mathbb{P}_q} &= \frac{\exp(\beta x_j)/\sum_{k=1}^{J}\exp(\beta x_k)}{\exp(\beta x_q)/\sum_{k=1}^{J}\exp(\beta x_k)} \\ &= \frac{\exp(\beta x_j)}{\exp(\beta x_q)},\end{aligned}$$

which is independent of any other alternative k

Independence of Irrelevant Alternatives

Logit errors imply for any two alternatives j and j'

$$\frac{\mathbb{P}_j}{\mathbb{P}_q} = \frac{\exp(\beta x_j)/\sum_{k=1}^{J}\exp(\beta x_k)}{\exp(\beta x_q)/\sum_{k=1}^{J}\exp(\beta x_k)} = \frac{\exp(\beta x_j)}{\exp(\beta x_q)},$$

which is independent of any other alternative k

This is called Independence of Irrelevant Alternatives

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

Logit errors also imply proportional substitution patterns:

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

$\Longrightarrow$ holds for all j.

In the case of the Red-bus-Blue-bus problem if p_{RedBus} goes up by 1%:

$$\%\text{ Change in } \mathbb{P}_{\text{Blue Bus}} = \%\text{ Change in } \mathbb{P}_{\text{Car}}$$

Model with correlated effects: Nested Logit

- Assume now that when choosing j agent i receives utility

$$V_{ij} = \beta x_j + \varepsilon_{ij}$$

with

$$\varepsilon_{ij} = \xi_{k(j)} + \epsilon_{ij},$$

where $k(j)$ denotes a category where product j belongs to.

- Now the idiosyncratic taste shock is correlated among products in the same category due to the presence of $\xi_{k(j)}$
- Categories of products, $k(j)$, are typically assumed known.
- In this work we will relax this assumption and estimate $k(j)$ together with the rest of the parameters of the model.

Model with correlated effects: Nested Logit (cont')

Recall the previous nested logit model

$$V_{ij} = \beta x_j + \varepsilon_{ij},$$

Assume $(\varepsilon_{i1}, ..., \varepsilon_{iJ})$ has cumulative distribution

$$\sim \exp\Big(- \sum_{k=1}^{K} \big(\sum_{j \in B_K} e^{-\frac{\epsilon_j}{\lambda_k}} \big)^{\lambda_k} \Big)$$

which gives closed-form solution probabilities

$$\begin{aligned}
\mathbb{P}_j &= \frac{e^{\frac{\beta x_j}{\lambda_k}} \Big(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \Big)^{\lambda_k - 1}}{\sum_{l=1}^{K} \Big(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \Big)^{\lambda_l}} \\
&= \frac{e^{\frac{\beta x_j}{\lambda_k}}}{\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}}} \frac{\Big(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \Big)^{\lambda_k}}{\sum_{l=1}^{K} \Big(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \Big)^{\lambda_l}} \\
&= \mathbb{P}_{j|k(j)} \mathbb{P}_{k(j)}
\end{aligned}$$

Nested Logit: Comments

- Nested Logit can be thought of as sequential choices across categories. Consumer choice: first nest, then alternative within nest

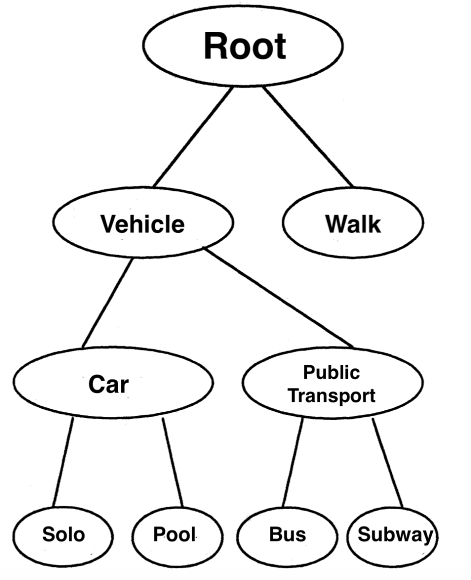
$$\mathbb{P}_j = \mathbb{P}(j|nest)\mathbb{P}(nest)$$

- $1 - \lambda_k$ can be interpreted as correlation within nest
- Products within categories are more substitutes than products across nests
- Recall for Logit, elasticities given by $E_j^q = -\beta p_q \mathbb{P}_q$
- For Nested Logit, elasticities given by

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

- Elasticity Multinomial Logit $\leq$ Elasticity Nested Logit

Example: Mean of Transport (Cardell, 1997)



Econometric Model

DGP

- We assume there is a large population of N individuals in a market choosing one out of J products
- Consideration sets are fixed to all products for all individuals
- Each product is chosen by a sufficient number of individuals so that:

$$\sup_{j \in \{1,2,\ldots,J\}} (\hat{\mathbb{P}}_j - \mathbb{P}_j) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

  where $\hat{\mathbb{P}}_j = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{\text{i buys j}\}$.

- We assume observable characteristics drive demand for products, and unobservables are grouped.
- We assume individuals take decisions consistent with a common nested logit model.
- Example: Demand for neighborhoods

Identification of groups (example in the case of 1 level)

Recall

$$\mathbb{P}_j = \frac{e^{\frac{\beta x_j}{\lambda_k}} \left(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \right)^{\lambda_l}}$$

Taking logs

$$\log \mathbb{P}_j = \beta_{k(j)} x_j + \xi_{k(j)},$$

where

$$\beta_{k(j)} = \frac{\beta}{\lambda_k} \qquad \text{and} \qquad \xi_{k(j)} = (\lambda_k - 1) \log \Big(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \Big) - \log \Big(\sum_{l=1}^{K} \Big(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \Big)^{\lambda_l} \Big)$$

Identification of groups (example in the case of 1 level)

Recall

$$\mathbb{P}_j = \frac{e^{\frac{\beta x_j}{\lambda_k}} \left(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \right)^{\lambda_l}}$$

Taking logs

$$\log \mathbb{P}_j = \beta_{k(j)} x_j + \xi_{k(j)},$$

where

$$\beta_{k(j)} = \frac{\beta}{\lambda_k} \quad \text{and} \quad \xi_{k(j)} = (\lambda_k - 1) \log \left(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \right) - \log \left(\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \right)^{\lambda_l} \right)$$

Conditional on x_j all parameters are only a function of group $k(j)$

Estimation

The estimation follows a 2-step procedure:

1. k-means given K number of groups using market shares and observables, x:
$$\underset{(k(1),k(2),\ldots,k(J);H_1,\ldots,H_K)}{\operatorname{argmin}} \frac{1}{J}\sum_{j=1}^{J} \left\|(\ln \hat{\mathbb{P}}_j, x_j) - H_{k(j)}\right\|^2$$
where $\hat{\mathbb{P}}_j$ is the sample counterpart of $\mathbb{P}_j$.
2. Conditional on the estimated groups, estimate parameters in nested logit model by maximum likelihood.

This estimator is the Grouped Fixed Effects estimator in Bonhomme, Lamadon, Manresa (2019).

Monte Carlo Simulations: DGP

- Number of products J, number of consumers N
- Draw product characteristics i.i.d. $x \sim \mathcal{N}(0, \sigma_x^2)$
- Set number of nests K
- Classify products using k-means on x
  $\Longrightarrow$ no correlation vs. correlation in product characteristics
- Draw correlation structure i.i.d $\lambda_k \sim U[0, 1]$
- Set $\beta = c$

Monte Carlo Simulations: DGP (cont)

- Construct population probabilities according to

$$\mathbb{P}_j = \frac{e^{\frac{\beta x_j}{\lambda_k}} \left(\sum_{d \in B_k} e^{\frac{\beta x_d}{\lambda_k}} \right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} e^{\frac{\beta x_d}{\lambda_l}} \right)^{\lambda_l}}$$

- Generate

$$\hat{\mathbb{P}}_j = \mathbb{P}_j + u_j, \qquad \text{where } u_j \sim \mathcal{N}(0, \frac{\mathbb{P}_j(1 - \mathbb{P}_j)}{N})$$

Monte Carlo Simulation: two-level nested logit

- Market size $= 10^6$
- Number of products $= 10^3$
- Just one characteristic $x \sim \mathcal{N}(0, 0.5)$. Set $\beta = 1$
- Group products according to x
- 2 upper nest, 6 lower nest
- Draw λ_U, λ_L from $U[0,1]$
- Construct $\ln \hat{\mathbb{P}}_j$
- Normalize outside option $x_J = 0, \lambda_{k(J)} = 1$, so

$$\ln \mathbb{P}_j - \ln \mathbb{P}_J = \beta_{k(j)} x_j + \xi^L_{k(j)} + \xi^U_{k(j)}$$

Recovering Parameters (1 estimation)

	Δ ln prob hat	
	True	Cluster
L_nest2	2.832***	2.874***
	(0.001)	(0.004)
L_nest3	8.924***	8.924***
	(0.0005)	(0.004)
L_nest4	9.567***	9.567***
	(0.0002)	(0.002)
L_nest5	8.750***	8.750***
	(0.0004)	(0.005)
L_nest6	3.201***	3.201***
	(0.0004)	(0.006)
L_nest1:x	7.164***	7.164***
	(0.001)	(0.008)
L_nest2:x	1.947***	1.908***
	(0.001)	(0.007)
L_nest3:x	1.222***	1.222***
	(0.001)	(0.007)
L_nest4:x	1.000***	1.000***
	(0.001)	(0.008)
L_nest5:x	10.869***	10.869***
	(0.001)	(0.005)
L_nest6:x	1.742***	1.742***
	(0.001)	(0.006)
Constant	−9.568***	−9.568***
	(0.0001)	(0.001)

Recovering the whole nest structure

Observe that in the previous table we have estimated

$$\hat{\beta}_k \quad \text{and} \quad \hat{\xi}_k$$

Recall the two-level nested logit equation

$$\ln \mathbb{P}_j - \ln \mathbb{P}_J = \beta_{k(j)} x_j + \xi^L_{k(j)} + \xi^U_{k(j)}$$

Hence

$$\hat{\beta}_k \text{ estimates } \beta_{L(k)} \quad \text{and} \quad \hat{\xi}_{L(k)} \text{ estimates } \xi^L_{L(k)} + \xi^U_{U(k)}$$

How do we separate $\xi^L_{L(k)}$ from $\xi^U_{U(k)}$ given $\hat{\xi}_{L(k)}$?

Recovering the whole nest structure (cont)

From the structural model, it follows

$$\xi_{L(k)} = (\lambda_{L(k)} - 1) \log\Big(\sum_{j \in L(k)} \exp(\frac{\beta}{\lambda_{L(k)}} x_j)\Big)$$

Observe that $\hat{\beta}_k$ estimates $\beta_{L(k)} = \frac{\beta}{\lambda_{L(k)}}$
Moreover, from the structural model

$$\lambda_k = \frac{\beta_{k(J)}}{\beta_k}$$

because we have normalized $\lambda_{k(J)} = 1$

$\implies$ can construct plug-in estimator of $\xi_{L(k)}$ and $\xi_{U(k)}$

$$\hat{\xi}_{L(k)} \quad \text{and} \quad \hat{\xi}_{U(k)}$$

To recover upper nest structure

k-means on $\hat{\xi}_{U(k)}$ with $k = 1, ..., K_L$

Recovering the whole nest structure: Results

- Number of simulations: 500
- Number of consumers $M = 10^6$, number of products $J = 10^3$
- Set $\beta = 1$
- Number of Upper Nests $K_U = 3$. Number of Lower nests $K_L = 9$
- Draw $x \sim \mathcal{N}(0, 1)$

Results:

90.3% match in Lower Nests and 79.2% match in Upper Nests

Ways Forward

- Endogeneity: η_j demand shocks correlated with price p_j
- Observed individual heterogeneity: x_{ij}
- Fuzzy Nests: different individuals might have different models of substitution
- Use the method in real data: Neighborhood choice in Amsterdam

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

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

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). This implies

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

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). This implies

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

Given IIA, we still have $\frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$. 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?

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). This implies

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

Given IIA, we still have $\frac{\mathbb{P}_c}{\mathbb{P}_{bb}} = 1$. 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