AME

Week 11: Multinomial Logit

Sophie Bindslev, November 2022

UNIVERSITY OF COPENHAGEN

Today's Plan

- Random Utility Model (recap)
- Multinomial Logit (MNL)
- Re-scaling
- Normalisation
- Partial Effects for MNL
- Odds Ratio
- Bootstrap Standard Errors
- Tastes in MNL
- Your time to shine!

Random Utility Model

- The Random Utility Model (RUM) serves as a framework which allows us to model N individuals' choices over J alternatives

$$u_{ij} = v_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I} \tag{1}$$

$$y_i = \underset{j \in \{1,2,\ldots,J\}}{argmax} \; u_{ij} \tag{2}$$

where i is the individual, j is the alternative, $y_i \in \{1, 2, \ldots, J\}$ is the alternative chosen by i, u_{ij} is i's utility from choosing alternative j. u_{ij} is composed of deterministic utility v_{ij} and stochastic utility ε_{ij}

RUM and Conditional and Multinomial Logit

- How we model deterministic, observable utility v_{ij} determines which model we estimate
- Conditional Logit: $v_{ij} = \mathbf{x}_j\boldsymbol{\beta}$
  - Individuals' choices depend on the **observable attributes of different alternatives** $\mathbf{x}_j$ e.g. if alternatives are cars then attributes include price, colour, whether the car is electric etc.
  - People with different characteristics e.g. income or household size are assumed to respond **in the same way** to e.g. price increases
- Multinomial Logit: $v_{ij} = \mathbf{x}_i\boldsymbol{\beta}_j$
  - Individuals' choices depend on **their observable characteristics** $\mathbf{x}_i$ e.g. a richer person may prefer Tesla
  - People with different characteristics e.g. income or household size can respond **differently** to different alternatives (e.g. cars)
- Combined: People with different characteristics e.g. income or household size are allowed to respond **differently** to different attributes of different alternatives (e.g. price and size of cars)

Multinomial Logit (CL)

- Today we will consider Multinomial Logit
- Choice probabilities are given by:

$$P(y_i = j \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_j)}{\sum_{h=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_h)} \tag{3}$$

- Therefore the log-likelihood function is:

$$\ell_i = \mathbf{x}_i \boldsymbol{\beta}_j - \log \left(\sum_{h=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_h) \right) \tag{4}$$

- The derivation can be found in Train (2009) ch. 3
- Like last time, if you use macro rather than micro data the estimated choice probabilities will be equal to market shares

Scaling for Numerical Stability

- Like last time we max-rescale to aid numerical precision (because $\exp(.)$ explodes for high values which can lead to overflow):

$$v_{ij} = \mathbf{x}_i \boldsymbol{\beta}_j - \max_{j \in \{1,2,\dots,J\}} \mathbf{x}_i \boldsymbol{\beta}_j \tag{5}$$

Normalisation & Identification

- We need to set $\boldsymbol{\beta}_0 = 0_{K\times 1}$. That is, we pick a "baseline" alternative against which all other alternatives are compared. We do this as only $J-1$ of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J-1}$ are identified
- Why can we only identify $J-1$ of the J different $\boldsymbol{\beta}_j$? Do we have the same (or a similar) problem in Conditional Logit?

Partial Effects

- The partial effects are given by:
  continuous

$$\frac{\partial P(y_i = j \mid \mathbf{x}_i)}{\partial x_{ik}} = p_{ij} \left(\beta_{jk} - \sum_{l=1}^{J} p_{il} \beta_{il} \right) \tag{6}$$

  and **discrete**

$$\delta_j^d(x_{ik}) = P(y_i = j \mid x_{ik} = 1) - P(y_i = j \mid x_{ik} = 0) \tag{7}$$

- **Interpretation:** How does the probability of choosing alternative j change when the characteristic k of chooser $i \in \{1, 2, \ldots, N\}$ changes by one unit?

Odds Ratio

- The odds ratio between alternatives j and h only depends on these (j, h) alternatives' attributes:

$$\frac{P(y_i = j \mid x_i)}{P(y_i = h \mid x_i)} = \frac{\frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_j)}{\sum_{k=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_j)}}{\frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_h)}{\sum_{k=1}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_h)}} = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_j)}{\exp(\mathbf{x}_i \boldsymbol{\beta}_h)} \tag{8}$$

- **Interpretation:** The odds ratio is the probability of choosing one alternative relative to the probability of choosing another alternative

Bootstrap: Inference on Partial Effects

- Bootstrap is an alternative to the delta method when computing the standard errors of measures of interest (e.g. partial effects, elasticities, the logsum etc.)
- To use this procedure:
  A) draw a bootstrap sample with replacement.
  B) estimate $\hat{\theta}$ on this sample and compute partial effects.
  C) Repeat for each bootstrap sample.
  D) Compute the standard deviation in estimated partial effects across bootstrap samples.
- You'll need a high number of boostrap samples for this to perform well. That might take quite a while to run...
- To resample our data we randomise the numbers in an index. `np.random.choice(N, size = N, replace = True)` can be used

Tastes in Multinomial Logit

- Is MNL also susceptible to IIA? Why/Why not?
- Like Conditional Logit, Multinomial Logit can only capture systematic taste variation, not random taste variation. Mixed Logit would be needed to capture random taste variation.

Your time to shine!

- Fill in `mlogit_ante.py` and solve the problem set
- The bootstrap question can take a while to run - don't lose faith