AME

Week 10: Conditional Logit

Sophie Bindslev, November 2022

UNIVERSITY OF COPENHAGEN

Today's Plan

- Random Utility Model
- Conditional Logit (CL)
- Scaling
- Partial Effects for CL
- Elasticities
- Additional Features of CL
- Tastes in CL
- 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

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

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

where i is the individual, j is the alternative, $y_i \in \{1, 2, \dots, J\}$ is the alternative chosen by i, u_{ij} is i's utility from choosing alternative j. u_{ij} consists 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 be less price sensitive or have a stronger preference for luxury cars, a bigger household may care more about car size and so on
  - People with different characteristics e.g. income or household size are allowed to respond **differently** to e.g. price increases

Conditional Logit (CL)

- Choice probabilities are given by:

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

- Therefore the log-likelihood function is:

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

  which is derived by maximising the choice probabilities for the chosen discrete choice for each i. $\mathbf{x}_{y_i}$ contains attributes of the alternative y_i chosen by i

- If you use macro rather than micro data the estimated choice probabilities will be equal to market shares

Scaling for Numerical Stability

- $\exp(.)$ explodes for high values which can lead to overflow. This leads to numerical instability which we seek to avoid by max-rescaling our deterministic utilities.
- Choice probabilities are unaffected by scaling (Why?)

$$P(\mathbf{y} = j \mid \mathbf{X}) = \frac{\exp(\mathbf{x}_j \boldsymbol{\beta})}{\sum_{h=1}^{J} \exp(\mathbf{x}_h \boldsymbol{\beta})} = \frac{\exp(\mathbf{x}_j \boldsymbol{\beta} - K)}{\sum_{h=1}^{J} \exp(\mathbf{x}_h \boldsymbol{\beta} - K)} \tag{5}$$

- Therefore, we can max-rescale:

$$v_{ij} = \mathbf{x}_j \boldsymbol{\beta} - \max_{j \in \{1,2,\ldots,J\}} \mathbf{x}_j \boldsymbol{\beta} \tag{6}$$

Partial Effects

- As in the binary setting discussed last week we cannot interpret β as the partial effects of $\mathbf{x}_j$ on the probability of j being chosen $p_{ij}(\mathbf{x}_i) = P(y_i = j \mid \mathbf{x}_i)$
- Instead, we consider the own- and cross-effects:

$$\frac{\partial p_j(\mathbf{X})}{\partial \mathbf{x}_{j,k}} = p_j(\mathbf{X})(1 - p_j(\mathbf{X}))\beta_k \tag{7}$$

$$\frac{\partial p_j(\mathbf{X})}{\partial \mathbf{x}_{h,k}} = -p_j(\mathbf{X})p_h(\mathbf{X})\beta_k \tag{8}$$

  where $k \in \{1, 2, \dots K\}$ is an attribute of the $j^{th} \in \{1, 2, \dots J\}$ alternative

- **Interpretation**: How does the probability of choosing alternative j change when the k^{th} attribute of alternative $h \in \{1, 2, \dots, J\}$ changes by one unit?

Elasticities

- We can compute own- and cross-price elasticities using choice probabilities $p_j(\mathbf{X})$ (which are $N \times 1$ vectors):

$$\epsilon_{jk} = \frac{\partial p_j(\mathbf{X})}{\partial \mathbf{x}_{h,k}} \frac{\mathbf{x}_{h,k}}{p_j(\mathbf{X})} = \frac{\partial p_j(\mathbf{X})}{p_j(\mathbf{X})} \Big/ \frac{\partial \mathbf{x}_{h,k}}{\mathbf{x}_{h,k}} \tag{9}$$

$$= \frac{\text{Rel. change in } p_j(\mathbf{X})}{\text{Rel. change in } \mathbf{x}_{h,k}} \tag{10}$$

- **Interpretation**:
  $\mid \epsilon_{ijk} \mid < 1$: inelastic, choice prob. responds less than attributes
  $\mid \epsilon_{ijk} \mid = 1$: unit elastic, choice prob. responds like attributes
  $\mid \epsilon_{ijk} \mid > 1$: elastic, choice prob. responds by more than attributes
- For **inference** on these you will need standard errors: use Delta Method (last week) or Bootstrap (next week)

Additional Features of CL

- **Identification**: In CL we cannot hope to estimate the intercept nor the variance of the stochastic component of utility. Why?
- **Interpretation** of β_k: marginal utility of attribute k. Means we can compute the marginal utility of income if we have a price variable (v useful)! Why is this marginal utility of income useful?
- **Welfare Measure**: Compensating Variation using the logsum

Tastes in Conditional Logit

- **IIA: independence of irrelevant alternatives.**
  What does IIA mean for choice probabilities? What does IIA imply for *substitution patterns across alternatives* following a change in one alternative's attributes and thereby its market share?
- **Systematic taste variation, not random taste variation can be captured by CL.**
  How can we see this in the model specification? (Hint: remember how we specified u_{ij})

Your time to shine!

- You'll be working with 3-dimensional arrays today. They are well worth getting comfortable with (also the type of data used in assignment 3)
- Take a look at `estimation.py` which allows you to compute maximum likelihood estimates and their variances
- Fill in `clogit_ante.py` and solve the problem set
- The "Conditional Logit" Note is helpful additional info