The Mixed Logit Model: Estimation

Lecture #8

Miguel I. Gómez,
Dyson School of Applied Economics and Management
Cornell University

Cornell University

Objectives

- Understand theoretical foundations the mixed logit model
- How to estimate a mixed logit model in Stata

Overview

- Discrete Choice and Mixed Logit Model
- Background
- Data
- Utility and choice probability
- Stata Commands
- Result
- More exercises
- Appendix
- References

Discrete Choice Model – Mixed Logit Model

- Daniel McFadden and others developed the mixed logit, or random parameters logit model to address the issues mentioned before (Unrealistic substitution pattern, Taste variation, etc.).

- McFadden and Train (2000) show that a mixed logit can approximate any logit model. In other words, all other logit models (there are many, although I did not talk about them in this class) are subsets or special cases of this most general specification.

Discrete Choice Model – Mixed Logit Model

Idea

- The fundamental idea behind the mixed logit model is that the utility from any one choice is no longer independent of any other choice, but rather correlated through the introduction of another stochastic component to utility.

- There are a number of ways that this can be motivated, depending on the application, but in economics, the most intuitive way is to allow the parameters of the utility function – the price response, or the marginal utility of product attributes – to be random variables.

Discrete Choice Model – Mixed Logit Model (Cont.)

Idea

- Because the responses are random, they can be correlated among choices.
- Choices that are more “similar” in terms of their attributes will induce a higher degree of correlation than those that are dissimilar
- The model estimates, and the associated elasticities, will indicate that similar products draw market share away from each other more than they do from dissimilar products
- This is exactly what we want, and it explains why the mixed logit model is perhaps the most popular method used in economic applications today.

Discrete Choice Model – Mixed Logit Model

- Development of the mixed logit model
- Until now, we considered the utility: $U_{nj}= \beta x_j + \varepsilon_{nj}$ (1)
- Now, consider the utility: $U_{nj}= \beta_n x_j + \varepsilon_{nj}$ (2)
- Note that the only difference in these specifications is that the response parameter now varies over decision makers, n. Note that we could also allow the constant terms or other attribute valuations, but let's keep it simple for now.
- In this model, the coefficients β_n are random, so are inherently unobservable to the researchers– just as the logit error terms are. Assume the taste parameters are distributed normally (many alternatives exist) so we can write:

$$\beta_n \sim N\left(\bar{\beta}, \sigma_\beta\right) \qquad (3)$$

Discrete Choice Model – Mixed Logit Model

- The response parameter has a mean component, and a variance over individuals in the dataset. Those two parameters can be estimated for the given attribute.

- Describe the distribution of β_n generally as $f(\beta)$. If the researchers were able to observe β_n, then the logit probability is given by the same formula that we saw before as he or she will only choose alternative i over $j \neq i$ if $u_{ni} > u_{nj}$.

- Therefore, conditional on observing β_n, we can write the logit probability as:

$$P_{ni} = \frac{\exp(\beta_n x_i)}{\sum_{j=1}^{J} exp(\beta_n x_j)} \qquad (4)$$

Discrete Choice Model – Mixed Logit Model

- Equation 4 is the usual logit expression. However, knowing that the researcher cannot observe β_n, the logit probability is found by taking the expectation over the distribution of heterogeneity of decision makers:

$$P_{ni} = \int \frac{\exp(\beta_n x_i)}{\sum_{j=1}^{J} exp\left(\beta_n x_j\right)} f(\beta) d\beta \qquad (5)$$

Discrete Choice Model – Mixed Logit Model

Estimation

- Equation 5 involves integrating over the distribution of β

- If there is more than one random parameter, the integral becomes a multiple integral.

- There is no closed-form solution for the mean utility in the mixed logit case. Therefore, estimation typically involves a simulation, which is called simulated maximum likelihood (SML).

Discrete Choice Model – Mixed Logit Model

The SML method

- Step 1: Take a random draw, r, from the distribution of β and calculate the implied probability according to equation 4.

- Step 2: Repeat the draw thousands of times and take the average over all draws. The average over all of these is an unbiased estimator of the true probability by construction. The average is found over all R draws using:

$$\hat{P}_{ni} = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp(\beta_n^r x_i)}{\sum_{j=1}^{J} exp(\beta_n^r x_j)} \quad (6)$$

- Step 3: Equation 6 is substituted into the likelihood function, and then it is maximized.

Discrete Choice Model – Mixed Logit Model (Cont.)

Substitution pattern

- The ratio of mixed logit probabilities, $\frac{P_{ni}}{P_{nj}}$ depends on all the data , including attributes of alternatives other than i or j.
    - Notice that the denominators of the logit formula are inside the integrals and therefore do not cancel.
- The elasticity is different for each alternative j.

Discrete Choice Model – Mixed Logit Model (Cont.)

Example

- Anglers’ choices of fishing sites
- Utility: $U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$

  n: angler

  j: fishing site

  t: time

  x consists of the following variables:

  Fish stock, Aesthetics rating, Trip cost, Indicator that the angler’s guide lists the site as a major fishing site, Number of campgrounds, Number of recreation access areas, Number of restricted species, Log of the size of the site

Discrete Choice Model – Mixed Logit Model

- The estimation result is given in table 2.

Table 2: Estimation result of mixed logit model			
Variable	Parameter	Estimate	t-ratio
Fish stock	Mean of ln(coefficient)	−2.876	−4.738
	Std. dev. of ln(coefficient)	1.016	4.113
Aesthetics	Mean of ln(coefficient)	−0.794	−3.467
	Std. dev. of ln(coefficient)	0.849	6.152
Total cost	Mean of ln(coefficient)	−2.402	−38.127
	Std. dev. of ln(coefficient)	0.801	10.269
Guide lists as major	Mean of ln(coefficient)	1.018	3.522
	Std. dev. of ln(coefficient)	2.195	6.236
Campgrounds	Mean of ln(coefficient)	0.116	0.359
	Std. dev. of ln(coefficient)	1.655	3.805
Access areas	Mean of ln(coefficient)	−0.950	−2.632
	Std. dev. of ln(coefficient)	1.888	5.379
Restricted species	Mean of ln(coefficient)	−0.499	−3.809
	Std. dev. of ln(coefficient)	0.899	5.482
Log(size)	Mean of ln(coefficient)	0.5018	4.690
SLL at convergence		−1,932.33	

- Most of the mean coefficients take plausible signs.

- The standard deviation of each random coefficient is highly significant, indicating that these coefficients do indeed vary in the population.

Exercise - Background

- We want to understand the effect of in-store promotion in a shampoo market.

- Consumers choose one of the following alternatives in each choice occasion.
  - Head & Shoulders
  - L'Oréal
  - Pantene
  - Outside option (e.g. non-purchase or purchase of other brands)

- There are 300 consumers and each of them has 10 choice occasions (e.g. choice task or shopping trip).

Background (Cont.)

- Utility function consists of the following attributes.
  - Brand-specific intercept
  - Price
  - In-store promotion
- When consumers do not purchase anything, they obtain just an error term from that option.
- The purpose of this estimation exercise is to estimate the effect of in-store promotion on consumers’ choice of shampoo brand.
- We incorporate consumer heterogeneity in price response, and estimate the mixed logit model.

Data

- Variables in L7_DATA1.csv
  - sn: Observation ID
  - hh: Consumer ID
  - time: Choice-occasion ID

  - price1: Price of Head & Shoulder
  - price2: Price of L'Oréal
  - price3: Price of Pantene

  - promo1: Takes 1 if Head & Shoulder is promoted, 0 otherwise
  - promo2: Takes 1 if L'Oréal is promoted, 0 otherwise
  - promo3: Takes 1 if Pantene is promoted, 0 otherwise

  - choice1: Takes 1 if Head & Shoulder is purchased, 0 otherwise
  - choice2: Takes 1 if L'Oréal is purchased, 0 otherwise
  - choice3: Takes 1 if Pantene is purchased, 0 otherwise
  - choice4: Takes 1 if any of the shampoo brands above are not purchased, 0 otherwise

Utility and Choice Probability

- Utility of consumer h from choosing shampoo brand j at choice situation t:

$$U_{hjt} = \alpha_j \ \ + \beta_h PRICE_{jt} \ + \gamma PROMO_{jt} + \varepsilon_{hjt},$$

$$j \in \{Head\ \&\ Shoulder, L'Oréal, Pantene\},$$

$$\varepsilon_{hjt} \sim iid\ EV \qquad (1)$$

$$U_{h4t} = \varepsilon_{h4t}, \varepsilon_{h4t} \sim iid\ EV \qquad (2)$$

- The logit probability for consumer h of choosing shampoo brand j at choice situation t:

$$P_{hjt} = \int \frac{\exp\left(\alpha_j \ +\beta_h PRICE_{jt} + \gamma PROMO_{jt}\right)}{\sum_{j=1}^{3} exp\left(\alpha_j \ +\beta_h PRICE_{jt} + \gamma PROMO_{jt}\right) + 1} f(\beta) d\beta \qquad (3)$$

Utility and Choice Probability (Cont.)

- Exercise 1
  - For simplicity, we assume there is no consumer heterogeneity in price response. Also, assume the coefficient estimates below, and calculate choice probability for each alternative of sn 1.
    - $\alpha_{Head\ \&\ Shoulder} = 1$
    - $\alpha_{L'Oréal} = 1.5$
    - $\alpha_{Pantene} = 2$
    - $\beta = -1$
    - $\gamma = 2$
- Exercise 2
  - How is the alternative-specific conditional logit model estimated?
- Exercise 3
  - How is the mixed logit model estimated?

Stata Commands

- Data loading

  insheet using C:\Users\mig7\Documents\L7_DATA1.csv

  - Specify the file path and file name

- Descriptive statistics

  sum sn hh time price1 price2 price3 promo1 promo2 promo3 choice1 choice2 choice3 choice4

  - Check if the data is uploaded correctly

- price4 and orini4 are set to be zero because deterministic part of the utility from choosing outside option is assumed to be zero.

  gen price4 = 0

  gen promo4 = 0

  - Recall the utility.

Stata Commands (Cont.)

- Change the data format from wide to long

  reshape long price promo choice, i(sn) j(alt)

  - The Stata command for the mixed logit model accepts only the long data format.
  - price1, price2, price3 and priced4 are combined to price and the subscript is recognized by the new variable, alt.
  - promo1, promo2, promo3 and promo4 are combined to promo and the subscript is recognized by the new variable, alt.
  - choice1, choice2, choice3 and choice4 are combined to choice and the subscript is recognized by the new variable, alt.
  - In the parenthesis of j, you need to specify the new variable, alt.
  - In the parenthesis of i, you need to specify the variable that specify the choice situation and is unique in the wide data.

Stata Commands (Cont.)

- Create alternative specific binary variables

```
gen alt1 = (alt == 1)   *If alt = 1, then alt1=1, zero otherwise.
gen alt2 = (alt == 2)   *If alt = 2, then alt2=1, zero otherwise.
gen alt3 = (alt == 3)   *If alt = 3, then alt3=1, zero otherwise.
```

Stata Commands (Cont.)

- Alternative-specific conditional logit

  asclogit choice alt1 alt2 alt3 promo price, case(sn) alternatives(alt) noconstant

  - choice : Dependent variable or choice
  - alt1 alt2 alt3 promo price: Independent variable(s) whose coefficient(s) is (are) nonrandom
  - case(·): Identifier variable for the choice occasions
  - alternatives(·): Identifier variable for choice alternatives

Stata Commands (Cont.)

- Mixed logit

```
global randvars "price"
mixlogit choice alt1 alt2 alt3 promo, rand($randvars) group(sn) id(hh)
nrep(100)
```

  - choice : Dependent variable or choice
  - alt1 alt2 alt3 promo: Independent variable(s) whose coefficient(s) is (are) nonrandom
  - Rand(·): Independent variable(s) whose coefficient(s) is (are) random
  - group(·): Identifier variable for the choice occasions
  - id(·): Identifier variable for the decision makers. This option is necessary when each individual performs several choices.
  - nrep(·): Number of Halton draws (number of random draws)

Results

Table 1: Result from the Simple and Mixed Logit Models

		Simple Logit		Mixed Logit	
	Variable	Estimate	t-ratio	Estimate	t-ratio
Nonrandom parameter	Head & Shoulder	0.48	2.44	1.08	5.14
	L'Oréal	0.92	5.11	1.53	7.86
	Pantene	1.46	7.77	2.08	10.28
	Promotion	2.06	29.77	2.11	29.54
Mean for random parameter	Price	−0.91	−4.58	−1.03	−4.69
Std. dev. for random parameter	Price	–	–	1.52	13.12
Log likelihood value		−3500.03		−3385.09	

Results (Cont.)

- Akaike Information Criterion (AIC)
  - $AIC = -2l(\hat{\beta}) + 2k$, where k is the number of parameters
  - The smaller is better.
    - $AIC_{MNL} = -2 \times (-3500.03) + 2 \times 5 = 7010.06$
    - $AIC_{MXL} = -2 \times (-3385.09) + 2 \times 6 = 6782.18$
    - AIC_{MXL} is the smaller.

- Bayesian Information Criterion (BIC)
  - $BIC = -2l(\hat{\beta}) + k \times logN$, where k is the number of parameters and N is the number of observations
  - BIC more penalize the increased number of parameters
  - The smaller is better.
    - $BIC_{MNL} = -2 \times (-3500.03) + 5 \times log3000 = 7040.09$
    - $BIC_{MXL} = -2 \times (-3385.09) + 6 \times log3000 = 6818.22$
    - BIC_{MXL} is the smaller.

Results (Cont.)

- Likelihood ratio test
  - Compare the likelihood of a null model with an alternative model.
  - The test statistic is: $LR = -2log\frac{L(\widehat{\beta_0})}{L(\hat{\beta})} = -2\{l(\widehat{\beta_0}) - l(\hat{\beta})\}$
    - $L(\widehat{\beta_0})$: likelihood value for a null model
    - $L(\hat{\beta})$: likelihood value for an alternative model
    - $l(\widehat{\beta_0})$: log-likelihood value for a null model
    - $l(\hat{\beta})$: log-likelihood value for an alternative model
  - $LR \sim \chi^2(g)$, where g is the difference in the number of parameters in the null and alternative model
  - If LR exceeds a critical value, then the null model is rejected.
    - Regard the asclogit model as a null model
    - $LR = -2 \times \{-3500.03 - (-3385.09)\} = 229.88$
    - $LR_{criti} = 7.88\ (g = 1)$
    - The asclogit model is rejected.

Result (Cont.)

Table 1: Result from the Simple and Mixed Logit Models

		Simple Logit		Mixed Logit	
	Variable	Estimate	t-ratio	Estimate	t-ratio
Nonrandom parameter	Head & Shoulder	0.48	2.44	1.08	5.14
	L'Oréal	0.92	5.11	1.53	7.86
	Pantene	1.46	7.77	2.08	10.28
	Promotion	2.06	29.77	2.11	29.54
Mean for random parameter	Price	−0.91	−4.58	−1.03	−4.69
Std. dev. for random parameter	Price	–	–	1.52	13.12
Log likelihood value		−3500.03		−3385.09	

- The mixed logit fits the data better.
- Pantene is most preferred
- Head & Shoulder is least preferred.
- Promotion has a positive impact on consumer choice.
- The law of demand holds as expected.
- Marginal utility from income (price coefficient) differs among consumers.

More Exercises

- Suppose you have a household panel scanner data set in the breakfast cereal category (L7_DATA2.csv).
- There are 3 brands: Cheerios, Fruit Loops and Special K, and an outside option.
- For each brand, the information about price (price1, price2 and price3) and sugar content (sugar1, sugar2 and sugar3) is available.

- Exercise 4
  - Use L7_DATA_Cereals.csv and find the best model to predict consumer purchase behaviors in the breakfast cereal market. The definition of each variable is same as before.

References

- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata.* Stata Press.
- Hole, A. R. (2007). Estimating mixed logit models using maximum simulated likelihood. *Stata Journal, 7(3),* 388–401.
- Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.