Demand Modeling - 1.202

Case Study 3: Multinomial Choice Models (MNL)

Due date: Friday April 12, 2019

Part 1: Developing a Multinomial Choice Model (40 points)

In this part, you are asked to estimate a multinomial choice model based on real world data using PandasBiogeme. You have the option to choose one dataset (out of two). Please refer to the "Discrete Choice Case Study.pdf" – Part 1: Multinomial Logit Models for the details. There you are provided with example model estimations and evaluations for each case study which you can use as your starting point.

Part 2: Forecasting Using the Multinomial Choice Model (30 points)

In this part, you are asked to forecast the effects of some policy scenarios using your best model developed in Part 1. Please refer to the "Discrete Choice Case Study.pdf" – Part 2: Forecasting for the details. There you are provided with example forecasting analyses for each case study which you can use as your starting point. You need to use PandasBiogeme and maybe a spreadsheet application for this part.

Part 3: Supplemental Problems (30 points)

Supplemental Problem 1 (10 points)

You are considering estimating a logit mode choice model for three different alternatives: carpool (1), drive alone (2), and public transit (3). Below are five different possible model specifications. For each specification, determine whether the coefficients of the model are in fact estimable, and explain how you arrived at your conclusion. It is advisable to come up with the specification tables as done in the class and recitation for an easier understanding.

Specification 1:

D_1 = 1 in carpool alternative
= 0 otherwise

D_2 = 1 in drive alone alternative
= 0 otherwise

TT = total travel time (for each alternative)

$DCITY_1$ = 1 in carpool alternative if person works in downtown
= 0 otherwise

$DSUBURB_1$ = 1 in carpool alternative if person doesn't work downtown
= 0 otherwise

Specification 2:

D_1 = as in Specification 1

D_2 = as in Specification 1

TT = as in Specification 1

$DCITY_1$ = as in Specification 1

$DCITY_2$ = 1 in drive alone alternative if person works downtown
= 0 otherwise

$DSUBURB_3$ = 1 in transit alternative if person doesn't work downtown
= 0 otherwise

Specification 3

Same as in Specification 2, but omitting $DCITY_1$

Specification 4

D_1 = as in Specification 2

D_2 = as in Specification 2

TT = as in Specification 2

$DCITY_1$ = as in Specification 2

$DCITY_2$ = as in Specification 2

A_1 = autos owned in carpool alternative
= 0 otherwise

A_2 = autos owned in drive alone alternative
= 0 otherwise

Specification 5

D_1 = as in Specification 2

D_2 = as in Specification 2

TT = as in Specification 2

$DCITY_1$ = as in Specification 2

$DCITY_2$ = as in Specification 2

A_3 = autos owned if person actually used transit in the transit alternative
= 0 otherwise

Supplemental Problem 2 (10 points)

Suppose we have the following information from a sample of 450 PC owners at MIT:

Type of PC	Number of Observations
IBM	97
IBM Clone	213
Macintosh	140

Given these limited data, you hypothesize that people choose PC's based on a simple multinomial model where the systematic utility of each alternative is a constant term; i.e.:

$$P_n(\text{type of PC} = i) = \frac{e^{\alpha_i}}{\sum_j e^{\alpha_j}}$$

1. Formulate the log-likelihood function for this model.
2. Determine analytically the maximum likelihood estimator for the α's and calculate these estimates empirically using the given data.
3. Estimate the asymptotic standard error of the estimates in question 2.

   Hint: Recall that Maximum Likelihood Estimators are asymptotically unbiased and asymptotically efficient. Asymptotic efficiency means that MLEs attain the Cramer-Rao Lower Bound $\mathrm{Var}(\hat{\alpha}) = -E\left[\nabla^2 L\right]^{-1}$

Supplemental Problem 3 (10 points)

In a linear city there are two stores located at points C and D as shown in the figure below. There are 500 daily trips, total, to both stores. Two hundred trips originate from the residential areas on the left of C, and 300 trips originate from the area to the right of D. (The area between C and D is not residential.)

The shoppers make one trip per day to one of the stores, either store 1 or store 2 but not both. Those who live on the left of point C are uniformly distributed between points A and B. The shoppers who live on the right of point D are exponentially distributed:

$$f_X(x) = \gamma \exp(-\gamma x), x > 0$$

where γ is an unknown constant (parameter) and x is the distance from point D (to the right). The utility to individual n from choosing store i (i = 1, 2) is given by:

$$U_{in} = \beta_1(\text{distance}_{in}) + \beta_2 \ln(\text{size}_i) + \varepsilon_{in}$$

where distance_{in} is the distance of individual n from store i, size_i is the size of store i, and ε_{in} is an error term, which is independently and identically distributed Gumbel (Type I Extreme Value) across all alternatives and across the population.

Given distances d_{AB}, d_{BC} and d_{CD}, what is the expected number of trips taken each day to store 1 and store 2? How does the value of γ affect the number of trips to stores 1 and 2?