Demand Modeling - 1.202

Case Study 3: Multinomial Choice Models (MNL)

Due date: Friday April 11, 2025

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) from “Data/”. Please refer to the “Discrete_Choice_Case_Study.pdf” – Part 1 (p.1) 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. Examples are available in “Examples/”.

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 (p.15) 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. Examples are available in “Examples/”.

Part 3: Supplemental Problems (30 points)

Problem 1 (15 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)
DCITY1	= 1 in carpool alternative if person works in downtown; otherwise
DSUBURB1	= 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

Problem 2 (15 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.:

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

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: $Var(\hat{\alpha}) \geq -E[\nabla^2 L]^{-1}$