Mathematical Modeling of Behavior (2022-2023)

Michel Bierlaire. TAs: Silvia Varotto, Cloe Cortes, Tom Haering, Nicola Ortelli

Course assignment — October 11, 2022

Aim and content

The goal of this assignment is to develop a discrete choice model for analyzing the transportation mode choice behavior of individuals living in the London metropolitan area. You will start by testing different model specifications so as to obtain the best model that you can. You will then use the resulting model to forecast the changes in demand in response to different changes in the attributes of the alternatives.

The description of the data can be found on Moodle under the *Datasets* link (London passenger mode choice, LPMC). Note that each group is given a different sample of the original LPMC data, available under the *Assignments* section on Moodle. All alternatives are available to all individuals.

Technical information

- All the necessary concepts to answer the problems at hand are covered during the lectures and/or laboratory sessions.
- The results of your analysis must be documented in a report, **written in English**. We provide a .tex template.
- Justify all your answers by explaining how the results were obtained and by reporting any used formula. Make sure that your answers are concise.
- **All members of the group must equally contribute in all tasks!**

Submission and grading

- Each group must submit **one** report as a .pdf file on Moodle under *Assignments*. Please name your report as `Group_XX.pdf`, where XX is your group number. Please submit only the report. If we need any additional file (e.g., any of the code associated with your model specifications) we will explicitly ask you to send it to us.
- The deadline to submit the report is **December 23, 2022 at noon**.
- This assignment accounts for **20% of your final grade**. **20 points** are distributed among the tasks described in the following section.

EPFL **Mathematical Modeling of Behavior (2022-2023)** TRANSP-OR

Michel Bierlaire. TAs: Silvia Varotto, Cloe Cortes, Tom Haering, Nicola Ortelli

Course assignment — October 11, 2022

Tasks

Model 0 [1 point] Start with a model specification that includes alternative specific constants, and cost and travel time of the different alternatives associated with generic parameters. Report both the specification (i.e., the utility functions) and the estimation results (parameter values, t-tests or p-values, null and final log likelihoods). **[0.5 point]**

1. Comment on the estimation output (sign and significance of all parameters). **[0.5 point]**

Model 1 [2.5 points] Using Model 0 as the base model, include alternative-specific parameters for one of the attributes of Model 0. Report both the specification and the estimation results (as defined previously). **[0.5 point]**

1. State the underlying assumption of defining alternative-specific parameters in this specific situation. **[0.5 point]**
2. Comment on the estimation output (as defined previously, including any changes from the previous model). **[0.5 point]**
3. Compare Model 0 and Model 1 with an appropriate statistical test. Justify your choice of test. State the null hypothesis and the result of the test. Denote the preferred model as $\text{Model}_{\text{pref}}$. **[1 point]**

Model 2 [3.5 points] Using $\text{Model}_{\text{pref}}$ as the base model, include one additional alternative attribute and one interaction of a socioeconomic characteristic with either the ASCs or one of the attributes. Report both the specification and the estimation results (as defined previously). **[0.5 point]**

1. State the underlying assumptions of the additional attribute and interaction in this specific situation. **[1.0 point]**
2. Comment on the estimation output (as defined previously). **[1.0 point]**
3. Compare $\text{Model}_{\text{pref}}$ and Model 2 with an appropriate statistical test. Justify your choice of test. State the null hypothesis and the result of the test. Denote the preferred model as $\text{Model}_{\text{pref}}$. **[1.0 point]**

Model 3 [2.5 points] Using $\text{Model}_{\text{pref}}$ as the base model, include an appropriate non-linear transformation of one of the variables. Report both the specification and the estimation results (as defined previously). **[0.5 point]**

1. State the underlying assumption of the non-linear specification defined in this situation. **[0.5 point]**
2. Comment on the estimation output (as defined previously). **[0.5 point]**
3. Compare $\text{Model}_{\text{pref}}$ and Model 3 with an appropriate statistical test. Justify your choice of test. State the null hypothesis and the result of the test. Denote the preferred model as $\text{Model}_{\text{pref}}$. **[1 point]**

Model 4 [3 points] Using $\text{Model}_{\text{pref}}$ as the base model, propose and test a nested or cross-nested structure. Report the nesting structure by means of a graph, together with the specification and the estimation results (as defined previously). **[1 point]**

1. State the underlying assumption of the proposed nesting structure in this specific situation. **[0.5 point]**
2. Comment on the estimation output (as defined previously). **[0.5 point]**
3. Compare $\text{Model}_{\text{pref}}$ and Model 4 with an appropriate statistical test. Justify your choice of test. State the null hypothesis and the result of the test. Denote the preferred model as $\text{Model}_{\text{pref}}$. **[1 point]**

Market shares [2.5 points] Assume that stratified random sampling was used to produce your sample. We consider the following strata:

- **S1:** females aged 40 years or younger;
- **S2:** females aged 41 years or older;
- **S3:** males aged 40 years or younger;
- **S4:** males aged 41 years or older.

	Age ≤ 40	Age > 40
Male	2'676'249	1'633'263
Female	2'599'058	1'765'143

Table 1: London population estimates in 2015 (Source: ONS)

Table 1 gives the size of each category in the full population.

1. Report the size and weight of each stratum in your sample [**1 point**].
2. Using $\text{Model}_{\text{pref}}$ and the weights of your strata, compute the predicted market share of each mode and their confidence intervals. Do the obtained results match your expectations? [**1 point**]
3. Compare the market shares predicted by $\text{Model}_{\text{pref}}$ with the weighted market shares computed using the actual choices. [**0.5 point**]

Forecasting [**5 points**] Consider the following scenarios: (i) an increase of car cost by 15%; and (ii) a decrease of public transport cost by 15%.

1. Report the market shares predicted by $\text{Model}_{\text{pref}}$ for each scenario. Do they match your expectations? Compare those with the original market shares. [**1 point**]
2. Which scenario is the most effective policy if the goal is to decrease the share of car? Explain why. [**0.5 point**]
3. Which scenario reports the highest public transportation total revenue? Explain why. Is it higher than the total revenue obtained without any of the policies? Can you explain why? [**0.5 point**]
4. Calculate the average value of time for car and public transportation. Comment on the obtained results. [**1 point**]
5. Compute the direct and cross aggregate elasticities of car cost and public transport cost and comment on the obtained results. Report the normalization factors. [**2 points**]