Project description: Semester project
Comparative analysis of discrete choice models estimation using different software packages

School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne

Student: Alexia Stéphanie Liviana Paratte Section of Mathematics

Professor: Michel Bierlaire
Supervisors: Evangelos Paschalidis, Negar Rezvany

1 Project description and motivation

Since their establishment in the field of econometrics, discrete choice models have been extensively used in many research fields such as transportation, environmental economics, health, energy, marketing, and others. Currently, there is a plethora of commercial and open-source software that can be used for the estimation of discrete choice models. Biogeme (Bierlaire, 2003) and Apollo (Hess & Palma, 2019) are two of the most appealing alternatives, as they are both open-source (based on Python and R respectively) and cover a broad range of advanced discrete choice models and diagnostics. More recently, other alternatives have been developed such as mixl (Molloy et al., 2021) and MO|DE.behave (Reul et al., 2023). Given the differences in the code development and the tools used, this project aims at investigating (a) differences in performance and computational speed between the two packages, (b) potential differences in the results and model statistics for the same models and datasets, and (c) available functionalities i.e., what are the common features and what are the specific features of some packages that are not available in others. Understanding the pros and cons of each package via this comparative analysis can ultimately lead to more optimised and accurate software for estimating discrete choice models.

2 Evaluation and learning objectives

The main objectives of this semester project are:

- **Literature review:** The student is expected to be familiar with the main concepts of the discrete choice modelling theory and the respective implementation in the software packages that will be used in the project (Biogeme and Apollo minimum).
- **Modelling:** The modelling process will follow the next steps:
  - Select (in agreement with the supervisors) the types of discrete choice models to be investigated.

  - Identify the respective available specifications in each of the software and code the equivalent specifications in the other software tools.
  - Evaluate and compare the results. (Model outputs, resource consumption, computational time etc.)
  - Discuss available functionalities i.e., what are the common features and what are the specific features of some packages that are not available in others.

- **Methods and data:** The student will use the data sets that are publicly available for each software. Templates from each software will also be used for model estimation. A proficiency level in Python language is required for this project, as Biogeme is based on Python. A good knowledge of other languages, such as R, is also desirable.

- **Report writing and software:** The student will write a clear and concise project report detailing all the model specifications, summarizing and critically discussing the results, highlighting the most significant findings.

3 Supervision

The project will be supervised by Evangelos Paschalidis, Negar Rezvany and Michel Bierlaire. The role of the supervisors is to provide guidance and feedback in all phases of the project. The student is advised to regularly write down the main findings from each phase of the project. The supervisor will validate the quality of the outputs for each project's phase and provide feedback on the work. If some of the results lead to the unexpected findings, the supervisor and student will re-define the new tasks accordingly. Regular meetings will be set up to discuss the progress of the project.

4 Project timeline and deadlines

This semester project will respect the following calendar:

- Start date of the project: 19.02.2024.
- Mid-term presentation date: to be defined
- End date of the project: 31.05.2024.
- Sending the report: to be defined
- Final presentation: to be defined
- Total number of weeks: 14 (spring semester 2024)
- Number of credits (ECTS): 10 (18-21 hours a week) [1 ECTS 25-30 hours]

Main deliverables:

- Midterm presentation: it should last 15-20 minutes and should include slides.
- Report: the final report and any developed code will be handed in one week before the final presentation.
- Final presentation: it should last 20 minutes and should include slides. presentation.

The expected deliverables of the semester is a Github repository with the code developed by the student and a written report. The code handed in must be sufficiently documented and well-structured. The report is expected to satisfy EPFL's requirements for both quality and plagiarism.

Signatures

_____Alexia Stéphanie Liviana Paratte

_____Michel Bierlaire

_____Paschalidis Evangelos

_____Negar Rezvany

Date : February 20, 2024 Place: Lausanne