SEMESTER PROJECT – MIDTERM PRESENTATION

PROJECT DESCRIPTION :

* Motivation :
  + Choice modelling is used to model the decision process of an individual or segment via revealed and/or stated preferences (revealed: observed actual behaviour, stated: hypothetical situation)
  + Widely used in transportation, for environmental issues, economics, marketing, etc. For now, will mainly focus on choices of transportation modes
  + Many tools on the market to do modelling. Will focus on two packages: *Biogeme* (in RStudio, developed by Bierlaire), and *Apollo* (in Python, developed by Hess & Palma).
  + The main goal is to compare the two packages. To do so, will look at the performances (computational speed, resources consumption), potential differences in the results and statistics (for same model/dataset), and features (how the packages differs or have the same features)
  + We want to try out different models to do so: multinomial logit model, mixed logit, etc.
* Process:
  + The process will always be as follows: select a model/dataset from Apollo website, translate it to Biogeme
  + Make sure the models are equivalent: to do so, I’ve read the documentation for both Apollo and Biogeme, to make sure to implement the script in Biogeme correctly. I had to look out at the ***.toml*** file in Biogeme and make sure the parameters works the same as in Apollo. We also need to look whether the **statistics** are the same (for example, the final loglikelihood). Also need to look whether it is **panel data** or not (Measurements over time. Panel data is a subset of longitudinal data where observations are for the same subjects each time)
  + Want to compare the estimation time and ideally the CPU and RAM usage in both cases

FIRST MODEL:

* Description:
  + First tried a basic Multinomial Logit Model on Apollo website -> get used to RStudio and the Apollo Package. It is a only revealed preference model, that I called RP\_Model.
  + The dataset is a synthetic one: mode choice for 500 travellers, which need to choose between car, bus, air or rail mode of transportation.
  + We need to estimate 9 variables coefficient: travel time (different for each mode of transportation), travel cost (same for all), access (same for everyone, car does not have one). Different availabilities for each individual (which I forgot to take into account at the beginning)
* Estimation results/Statistics
  + The parameters estimated are thus the ASC’s (the one for the car is set to 0), the Beta’s for the time, the beta for the access and the beta for the costs. We see that these are all almost equal
  + This is confirmed when looking at the final logelikehood, which is the same in both cases. This validates that the models are equivalent
  + However: too simple of a model, the computational time is almost the same, need a more complex model

SECOND MODEL:

* Description:
  + We choose an MMNL Model (Multinomial Mixture logit model), called MMNL\_Preference\_Model.
  + The dataset comes from a stated preference survey conducted in Switzerland In 2008. 388 people are faced with two alternatives of transportation routes. In this case, it is panel data, and they are not particular availabilities (everything is accessible for everyone)
  + Again, they are 9 parameters, but those are not directly estimated. We generated random coefficients using the Halton Draws, which is one of the two type of draws that are available in Apollo. Then, we estimate the means and standard deviations of these random coefficients
* All MMNL Models:
  + First one: Forgot about Panel Data, used the same starting values in Apollo and Biogeme. However, got very far results, which made us thought that biogeme was converging to a local solution
  + Second one: Still forgot about Panel Data, but to avoid the same problem as in the first one, we decided to change the starting value in Biogeme. To converge to the right local solution, we used the estimated parameters obtained in Apollo as starting values in Biogeme. Still, far results
  + Final one: This time, I took into account Panel Data
* Estimation Results/Statistics:
  + Still very far results, and I am not sure why. Estimation time are very different, but not sure we can compare as we get different estimations results and very far statistics.
  + Can also look at the CPU/RAM and see that the RAM is being used kind of the same, however the CPU is much more solicited that the RAM. Not sure what the peak correspond to
  + Will talk in the conclusion how we plan to avoid this problem

FUTURE GOALS

* Challenges faced:
  + Getting accustomed to Apollo was the first challenge. However, writing the script is kind of the same as in Biogeme
  + Did quite a lot of errors in translation from Apollo to Biogeme. Forgetting about Panel Data, had some problems with the NA values
  + Sometimes hard to find documentation in Biogeme. For example, I had been stuck with the *BioDraw* function in Biogeme, which makes it possible to choose which type of draws we want for the Mixture Model. However, I have finally found a document on it.
  + The CPU/RAM monitoring implementation was a bit hard as well. For now, I have coded it in Python, and launched it in a separated script as the codes for the Model. It is not directly implemented in the code. This would be nice to do in the future.
* Planning for the following weeks:
  + For the problem for the MMNL Model, we want to try something different than what we did until now. Want to choose a model and dataset from Biogeme this time, and translate it to Apollo. I am currently working on another Mixture Model chosen from the Biogeme Website
  + Another solution we considered it just to simplify the MMNL Model
  + After these equivalences issues solved, we want to dive deeper into the optimisation aspect. To do so, the lead we want to follow for now is to modify the parameters of the *.toml* file in Biogeme. The two parameters we want to look at are: 1. **The second derivatives** 2. The **optimization algorithm**.
  + We also want to look at different choice model that the simple Multinomial Logit Model and the Mixture model.