Semester Project - Week 2 Report

Alexia Paratte

March 21, 2024

Goals

After having the first meeting on the 29/07/2024, these are the goals set for the second meeting:

- Choose an MMNL (Mixture Model) on the Apollo website and estimate it

- Translate to Biogeme and insure the two models are equivalent

- Put the `monitor_system()` function directly into the code for Biogeme

- Translate `monitor_system` into RStudio and do the measurement

- Change the time measurement just before and after model estimation

- Look at the second derivative (Hessian) in `biogeme.toml`, which can go from 100 % to 0. Try to change the parameter: use 100%, 0% and some in between

- Change the method of estimating by modifying the file `biogeme.toml`. Need to change `optimization_algorithm` from `simple_bounds` (default) to `simple_bounds_BFGS`. Need to look at the Biogeme documentation to see exactly which one to choose

Intro

The first week consisted in coding a simple MNL Model for revealed preferences, and measure the performances (estimation time, CPU and RAM used) for both Biogeme and *Apollo*. However, the model was "too simple" to clearly see the difference of performances between the two languages (fast computation time in both cases). We thus decide to try again but with a MMNL model called `MMNL_Preference_Model`, which should take more time to estimate. It is a mixture of logit, using Monte-Carlo integration with Halton draws. The mixing distribution is a normal distribution.

Data Description

The dataset `apollo_swissRouteChoiceData` comes from an actual SP survey of public transport route choice conducted in Switzerland (Axhausen et al., 2008). A set of 388 people were faced with 9 choices each between two public transport routes, both using train (leading to 3'492 observations in the data). The two alternatives are described on the basis of travel time (tt), travel cost (tc), headway (hw) (time between subsequent trains/busses) and the number of interchanges (ch). For each individual, the dataset additionally contains information on income, car availability in the household, and whether the journey was made for commuting, shopping, business or leisure.

Models

The choice model is a Multinomial Logit Model (MMNL). We are trying two slightly different models, for reasons we will see later on.

1. MMNL Preference Model 0

Model parameters: For $i \in \{$tt, tc, hw, ch$\}$:

- $\mu_{log(\beta_i)}$ is estimated mean for the log of β_i, set to -3 before estimation

- $\sigma_{log(\beta_i)}$ is standard deviation of β_i, set to -0.01 before estimation

Utility functions:

$$V_{\text{alt1}} = \beta_{\text{tt}} \cdot \text{tt}_1 + \beta_{\text{tc}} \cdot \text{tc}_1 + \beta_{\text{hw}} \cdot \text{hw}_1 + \beta_{\text{ch}} \cdot \text{ch}_1$$
$$V_{\text{alt2}} = \beta_{\text{tt}} \cdot \text{tt}_2 + \beta_{\text{tc}} \cdot \text{tc}_2 + \beta_{\text{hw}} \cdot \text{hw}_2 + \beta_{\text{ch}} \cdot \text{ch}_2$$

where $\beta_i = -exp(\mu_{log(\beta_i)} + \sigma_{log(\beta_i)} \cdot \text{draws}_i)$ parameter for $i \in \{$tt, tc, hw, ch$\}$. These are the **random coefficients**. The randomness is brought by the draws$_i$ variables (. We generate them using the Halton Draws, which is often used in Monte-Carlo simulation. We choose to do **500** individual draws for both Biogeme and Apollo. In the first case, we replace the `biogeme.toml` by `few_draws.toml`, where we set [MonteCarlo] `number_of_draws` = 500. In the *Apollo* case, we set `interNDraws`=500 in the code directly.

Estimation:
We thus estimate the `MMNL_Preference_Model_0` in both Apollo and Biogeme. We see the results in the Table 1. We observe that the results for both languages are very different. This might be due to the fact that the number of parameters estimated in the Apollo case is 8. In Biogeme however, this number is set to 7. This is surprising, because when defining the 8 different parameters in both languages, we explicitly ask for all of them to be estimated.

Nevertheless, Biogeme seems to set and keep $\mu_{log(\beta_{ch})}$ to 0. To counter this, we decide to slightly modify this model.

Parameter	Value in Apollo	Value in Biogeme
$\mu_{log(\beta_{tt})}$	-1.984	-0.623
$\sigma_{log(\beta_{tt})}$	-0.442	-1.350
$\mu_{log(\beta_{tc})}$	-1.016	0.297
$\sigma_{log(\beta_{tc})}$	-0.991	-3.780
$\mu_{log(\beta_{hw})}$	-2.938	-1.690
$\sigma_{log(\beta_{hw})}$	-0.834	-2.150
$\mu_{log(\beta_{ch})}$	0.631	0
$\sigma_{log(\beta_{ch})}$	0.858	-5.540

Table 1: Parameter values for te MMNL_Preference_Model_0 in both Apollo and Biogeme librairies.

2. MMNL Preference Model 1

Model parameters:
Same as Model 0, however we set the starting values in the Biogeme code to the one found in Apollo. We thought Biogeme might want to find a local solution, so starting at the values found with Apollo might solve the problem.

Utility functions:

$$V_{alt1} = asc + \beta_{tt} \cdot tt_1 + \beta_{tc} \cdot tc_1 + \beta_{hw} \cdot hw_1 + \beta_{ch} \cdot ch_1$$
$$V_{alt2} = \beta_{tt} \cdot tt_2 + \beta_{tc} \cdot tc_2 + \beta_{hw} \cdot hw_2 + \beta_{ch} \cdot ch_2$$

We are adding in this case a constant asc, which is set to 0 and not estimated. We hope that, in this case, Biogeme will estimate $\mu_{log(\beta_{ch})}$.

Estimation:
As observed in Table 2, it seems that Biogeme did not want to estimate $\mu_{log(\beta_{ch})}$ and left it to 0. The number of estimated parameters is still 7. Furthermore, the estimated parameters in Biogeme are very far from the ones in Apollo, and the statistics observed in the HTLM of Biogeme does not make sense.

Optimization

Did not really started on this as the models do not seem equivalent. However, we can already see that the estimation time is not the same. For 500 draws in `MMNL_Preference_Model_1`, we get estimation times of **00:06:51** and **00:03:03** in Biogeme and Apollo respectively.

Parameter	Value in Apollo	Value in Biogeme
$\mu_{log(\beta_{tt})}$	-1.984	-25
$\sigma_{log(\beta_{tt})}$	-0.442	-26.1
$\mu_{log(\beta_{tc})}$	-1.016	-24.7
$\sigma_{log(\beta_{tc})}$	-0.991	-27.9
$\mu_{log(\beta_{hw})}$	-2.938	22.2
$\sigma_{log(\beta_{hw})}$	-0.834	-26.8
$\mu_{log(\beta_{ch})}$	0.631	0
$\sigma_{log(\beta_{ch})}$	0.858	28

Table 2: Parameter values for te MMNL_Preference_Model_1 in both Apollo and Biogeme librairies.

Questioning/Remarks

1. Cannot find information on the `bioDraws` function

2. Maybe we should have one ASC for each alternative, set one to 0 and estimate the other.

3. Need to try a bigger number of draws when the two models are equivalent, as to mesure performance better

4. `monitor_function`: I have tried to implement it in the code directly but not sure how to make it work as the estimation is running (it runs *then* makes the estimation...)