Group 17 – Assignment 3
Marija Nikolic
Yoann Trellu
Ma Yuanfang

Switzerland Mode Choice - Optima

Tests of the Model Structure - McFadden IIA Test

In the Optima data set we have the choice between three alternatives: public transports (PT), private modes (CAR); soft modes (MD). It is possible that there are common unobserved attributes between the two motorized alternatives (public transport and private modes). They might be correlated. In order to test this assumption, we perform the McFadden IIA test. In this test, we are interested in the value of the t - statistic for the coefficient related to the auxiliary variables BETA_IIA. As we can see in Table 1 BETA_IIA is significantly different from 0 at a 95% level of confidence, indicating that the IIA property does not hold for alternatives public transport and private modes. It means that those alternatives share some unobserved attributes. This is the reason why we decided to use Nested logit specification for our model which captures correlation between alternatives sharing some common characteristics.

Parameter name	Value	Standard error	t-test	p-value
BETA_IIA	4.29	0.518	8.27	0

Table 1. Estimation result for BETA_IIA parameter

The model specification is given in NL_model_optima.mod file, and estimation results are given in NL_model_optima.xml.

Prediction Tests - Outlier analysis

For our model, specified in NL_model_optima.mod file, and for all the observations in the optimaTOT3_valid.dat file, we calculate the predicted choice probability of the chosen alternative using Biosim. The distribution of these values is plotted in Figure 1. The distribution is shifted on the right meaning that the model predicts lots of high probabilities for the chosen alternatives. For checking outliers, we focus on the left part of the graph, which correspond to low predicted choice probabilities.

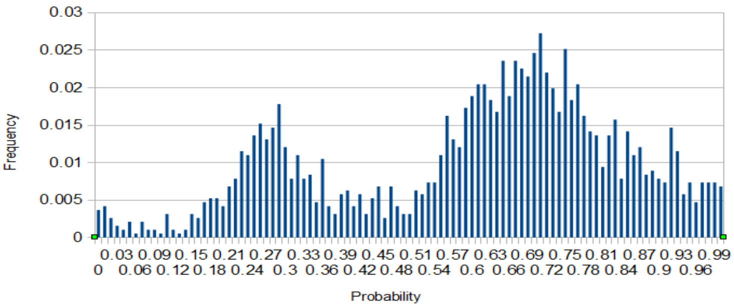

Figure 1.Outlier analysis

We checked for large deviations by inspecting all the observations with predicted probabilities less than some arbitrary small value, 0.01 in our case. What we found was that value of cost attribute of alternative which is not chosen was underestimated by decision-maker (for example cost for CAR was reported as 46 and cost for PT was reported as 230, and choice was PT – people tend to under/over estimate the attributes of alternative that they do not use). In transport mode choice models cost and time are the most important variables, therefore mentioned low predicted probabilities are not surprise.

Prediction Tests - Validation

The validation consists in studying the prediction of a model on external data. We have decided to divide the original data in two data sets, which contain half of the observations. One data set was used for estimation and another for validation purpose. Comparison for predicted and observed shares, using validation data set, is done for different market segments, based on income level. Results can be seen in Figure 2.

	Market share (predicted)	Market share (observed)
CAR	65.94%	65.90%
PT	28.08%	28.12%
MD	5.98%	5.98%
Total	100.00%	100.00%

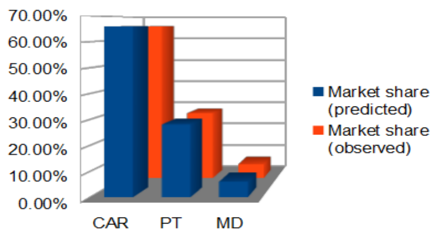

Figure 2. Validation

Note that values in the columns Market share (predicted) and Market share (observed) are mostly the same. This is logical because the model contains alternative specific constants (ASC_PT, ASC_CAR and ASC_MD). Conclusion is that the model is able to reproduce exactly the shares of the data.

Forecasting - Market shares

In the Figure 3 the current market shares for each of the three models are presented. On the left side you can see market shares computes using a segmentation based on income groups (low: inc=1,2 medium: inc=3,4 high: inc=5,6 and unknown income group).

	Low income (inc level = 1,2)	Medium income (inc level = 3,4)	High income (inc level = 5,6)	Unknown income
CAR	72.59%	62.49%	66.41%	74.16%
PT	20.91%	31.50%	27.59%	20.62%
MD	6.51%	6.01%	6.00%	5.21%
Total	100.00%	100.00%	100.00%	100.00%

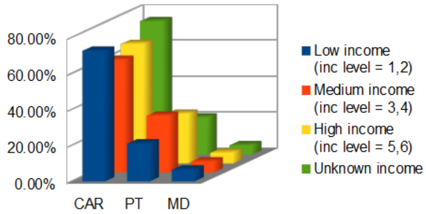

Figure 3. Market shares using segmentation based on income groups– base case

As we know, the purpose of developing a model is forecasting, and here we are interested in effect on market share of changes in one attribute. The influence of an increase of 35% of the cost for public transport is tested. The predicted shares by alternatives are presented in Table 2.

	Market share (Base case)	Market share (PT cost increase 35%)
CAR	65.94%	69.21%
PT	28.08%	24.80%
MD	5.98%	5.99%

Table 2. Overall market shares – Base case vs. PT cost increase 35%

When we look at the market shares in Table 2, we can see a decrease of the number of people choosing the alternative PT, which is logical as its cost has increased. The decrease is translated in an increase of the shares associated with the alternatives CAR and MD. The larger increase can be noticed for CAR than for MD alternative.

Note: We think that 35% is a big shift, and a decrease of 4% is not that much. Explanation can be that maybe people taking public transports put the priority on comfort, time, convenience, etc.

How standard errors of parameters can affect the predicted market shares?

For computing the market shares, the estimated parameters of the model are used (we do not know the real value). We should not blindly trust our model, because we put a lot of assumptions into it. Therefore it is necessary to check the results, which means to perform sensitivity analysis on BETA parameter of interest. The values of the BETA parameter have to be chosen from 95% confidence interval [BETA – 1.96*StdError, BETA + 1.96*StdError].

Price optimization - the optimal cost for public transport

The public transportation revenues can be calculated through multiplying public transportation cost by market shares of public transportation. The public transportation cost is estimated as the average of Margin Cost, which is equal to 11.1054. The market shares of public transportation and public transportation revenues according to percentage changes in public transportation costs are listed in Table 3. As PT_cost increases, the market share of public transportation decreases but the corresponding public transportation costs increase. As a result, the market share of Car increases and the market share of MD remains around 6%. It means that the higher the price of public transportation, the higher the public transportation revenues will be. In this case, there is no optimal value of PT_cost to get the highest public transportation revenues. The reason for that might be the specification of our model – it should be improved, validated and revenues calculated again.

	Market Shares Car	Market Shares MD	Market Shares PT	PT Revenues
PT_cost *3	77.02%	6.03%	16.94%	5.64
PT_cost *2	73.33%	6.01%	20.64%	4.58
PT_cost *1.8	72.27%	6.01%	21.71%	4.34
PT_cost *1.6	71.03%	6.00%	22.95%	4.08
PT_cost *1.4	69.59%	6.00%	24.40%	3.79
PT_cost *1.2	67.90%	5.99%	26.10%	3.48
PT_cost *1	65.94%	5.98%	28.08%	3.12
PT_cost *0.8	63.65%	5.98%	30.36%	2.70

Table 3. Public transportation revenues

Strategy to deal with the oversampling of men

We know that in Switzerland there are 97.3 men for 100 women, and we would like this ratio to be respected in the sample we are studying. We have at our disposition 943 male and 871 female respondents, obviously this is not the ratio we are looking for. This is why we will use a strategy based on weights to virtually work on a good calibrated sample. Weights in our case will logically give more importance to a single female respondent than a single male respondent. We will therefore look for values Wmale and Wfemale such that

$$\frac{(Wmale \times 943)}{(Wfemale \times 871)} = \frac{97.3}{100} \quad \text{and} \quad Wmale \times 943 + Wfemale \times 871 = 943 + 871$$

The first condition is about getting the right ratio, the second one is about the sample size: the weighted sample should be of same size. Hence, resolving this simple system give us Wmale = 0.948 and Wfemale = 1.056. Market shares based on gender market segmentation can be seen in Table 4.

Alternative	Predicted market share
CAR	65.18%
PT	27,56%
MD	5.93%

Table 4. Predicted market share

There is no big differences here compared to Table 2, but this makes sense since the weights found are still close to one.