Demand Modeling: 1.202
Spring 2019
Alexandra Berke

Case Study 4 - Report

This report is about the Swissmetro case.

Part 1

This work builds off of the work completed in Case Study 3. The best linear logit model found in that case study (see *SpecTest_SM_base.**) is specified as follows:

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

The log likelihood value for this specification is -4913.23. See appendix for estimated coefficient values and estimation results (Appendix Table 1).

The new models described in this report build on this one to test additional hypotheses and specifications.

1. **Market segmentation: Income**
2. **Non-nested hypotheses: income vs gender**
3. **Non-linear travel time**

1. Market Segmentation: Income

Files:
*SpecTest_SM_base.**
*SpecTest_SM_low_income.**
*SpecTest_SM_medium_income.**
*SpecTest_SM_high_income.**

Description

In this analysis, I hypothesize that travellers' preferences of whether to choose car, train, or SwissMetro, are affected by their income levels.

The population is segmented into groups depending on income, with the hypothesis that there is variation in preference across income levels. The reasoning for this segmentation is that people may have similar preferences when cost and income are not factors, but may not be able to afford those preferences due to their income level.

The market segments are as follows:

- Low income: All observations where the traveler's income is less than 50k CHF (variable INCOME=1)
- Medium income: All observations where the traveler's income is between 50k and 100k CHF (variable INCOME=2)
- High income: All observations where the traveler's income is more than 100k CHF (variable INCOME=3)

The null hypothesis is that there is no preference variation across income level market segments.

H_0: $B_{low\_income} = B_{medium\_income} = B_{high\_income}$

The model is estimated with the same specification for the full population as well as the income level segments, where the observations are excluded for the income level segments that do not have the corresponding income level.

Model specification:

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

The number of coefficients is 12.

Results:

Model	Log Likelihood	Number of coefficients
Restricted model (full population)	-4913.23	12
Low income	-668.7424	12
Medium income	-1562.482	12
High income	-1873.442	12

The log likelihood test with (12 + 12 + 12 - 12 = 24) 24 degrees of freedom yields:

$$LR = -2*(L_N(\beta) - \boldsymbol{\Sigma}\, L_{Ng}(\beta_g))$$
$$= -2*(-4913.23 + 668.7424 + 1562.48 + 1873.442) = 1617.1312$$

$X^2_{0.95,24} = 36.415 < 1617.1312$

We can therefore reject the null hypothesis at a 95% level of confidence. This validates the hypothesis that market segmentation occurs based on a traveller's income level.

2. Non-nested hypotheses: income vs gender

Files:
*SpecTest_SM_M1.**
*SpecTest_SM_M2.**
*SpecTest_SM_MC.**

Description
This analysis builds off of the previous results that there is significant market segmentation across income levels, as well as the finding that there is significant market segmentation among male versus female travelers (as shown in example).
These two findings may be related, or not. For this analysis, I compare three model specifications.
Model 1 adds dummy variables for low income and high income in order to account for the income market segmentation (medium income is normalized to zero).

Model 2 adds a dummy variable for male travelers to account for the significant difference in preferences of male vs female travelers.

Model C is a composite model of model 1 and model 2. Model 1 and model 2 are not nested, but both are restrictions of model C which has dummy variables for low income, high income, and male.

The added dummy variables have alternative specific coefficients because the hypothesis following the market segmentation analyses was that these variables affect travelers' preferences for the different travel mode options.

All models build off of the previous (*SpecTest_SM_base)* model. They are specified as follows:

Model 1 (M1):

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR +
B_car_low_income*LOW_INCOME + B_car_high_income*HIGH_INCOME

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST + B_train_low_income*LOW_INCOME +
B_train_high_income*HIGH_INCOME

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST + B_sm_low_income*LOW_INCOME +
B_sm_high_income*HIGH_INCOME
```

Model 2 (M2):

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR +
B_car_male*MALE

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST + B_train_male*MALE

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST + B_sm_male*MALE
```

Composite model (MC):

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR +
B_car_low_income*LOW_INCOME + B_car_high_income*HIGH_INCOME + B_car_male*MALE

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST + B_train_low_income*LOW_INCOME +
B_train_high_income*HIGH_INCOME + B_train_male*MALE

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST + B_sm_low_income*LOW_INCOME +
B_sm_high_income*HIGH_INCOME + B_sm_male*MALE
```

Where the dummy variables are defined as follows:

- LOW_INCOME has the value of 1 when the traveller has income of less than or equal 50k CHF, and 0 otherwise.
- HIGH_INCOME has the value of 1 when the traveller has income of greater than 100k CHF, and 0 otherwise.
- MALE has the value of 1 when the traveler is male.

Results

Model 1 estimation results:

Init log likelihood:	-6958.425
Final log likelihood:	-4851.029
Likelihood ratio test for the init. model:	4214.792
Rho-square for the init. model:	0.303

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.194	0.158	-1.23	0.218	0.17	-1.14	0.254
ASC_SM	0.213	0.128	1.67	0.0957	0.129	1.65	0.0993
B_CAR_COST	-0.00743	0.00107	-6.92	4.48e-12	0.0015	-4.94	7.64e-07
B_CAR_HIGH_INCOME	0.242	0.0542	4.47	7.93e-06	0.0533	4.55	5.46e-06
B_CAR_LOW_INCOME	0.0566	0.092	0.615	0.539	0.0891	0.635	0.525
B_CAR_TIME	-0.013	0.000811	-16	0	0.00164	-7.91	2.66e-15
B_FIRST	0.254	0.078	3.26	0.00113	0.0817	3.11	0.00188
B_GA	0.581	0.189	3.08	0.0021	0.196	2.96	0.00304
B_HE	-0.00633	0.00107	-5.94	2.93e-09	0.00108	-5.88	4.08e-09
B_SENIOR	-1.76	0.119	-14.8	0	0.11	-16	0
B_SM_COST	-0.0118	0.000617	-19.1	0	0.000927	-12.7	0
B_SM_HIGH_INCOME	0.248	0.0447	5.54	3.07e-08	0.0445	5.57	2.54e-08
B_SM_LOW_INCOME	-0.307	0.0617	-4.98	6.34e-07	0.0619	-4.96	6.97e-07
B_SM_TIME	-0.0108	0.000878	-12.3	0	0.00182	-5.94	2.86e-09
B_TRAIN_COST	-0.027	0.00128	-21.1	0	0.0021	-12.8	0
B_TRAIN_HIGH_INCOME	-0.49	0.0707	-6.93	4.17e-12	0.0705	-6.95	3.57e-12
B_TRAIN_LOW_INCOME	0.251	0.0739	3.39	0.000693	0.0738	3.39	0.000687
B_TRAIN_TIME	-0.00819	0.000847	-9.67	0	0.00118	-6.94	3.81e-12

Table 1: Estimation results from Model 1 (M1)

Model 2 estimation results:

Init log likelihood:	-6958.425
Final log likelihood:	-4828.744
Likelihood ratio test for the init. model:	4259.361
Rho-square for the init. model:	0.306

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-1.07	0.171	-6.3	3.04e-10	0.182	-5.91	3.46e-09
ASC_SM	-0.437	0.129	-3.38	0.00073	0.131	-3.34	0.000833
B_CAR_COST	-0.00806	0.00108	-7.46	8.86e-14	0.00151	-5.35	8.94e-08
B_CAR_MALE	0.609	0.0681	8.94	0	0.0714	8.53	0
B_CAR_TIME	-0.0131	0.000816	-16.1	0	0.00166	-7.91	2.44e-15
B_FIRST	0.339	0.0746	4.54	5.65e-06	0.0793	4.27	1.95e-05
B_GA	0.608	0.189	3.22	0.00128	0.202	3.01	0.00264
B_HE	-0.0063	0.00107	-5.9	3.61e-09	0.00108	-5.86	4.73e-09
B_SENIOR	-1.8	0.119	-15	0	0.111	-16.2	0
B_SM_COST	-0.0117	0.000616	-19	0	0.000925	-12.7	0
B_SM_MALE	0.187	0.0472	3.97	7.08e-05	0.0486	3.86	0.000114
B_SM_TIME	-0.0111	0.000882	-12.6	0	0.00184	-6.06	1.35e-09
B_TRAIN_COST	-0.028	0.00127	-22	0	0.00207	-13.5	0
B_TRAIN_MALE	-0.796	0.061	-13.1	0	0.0619	-12.9	0
B_TRAIN_TIME	-0.00863	0.00086	-10	0	0.00122	-7.1	1.24e-12

Table 2: Estimation results from Model 2 (M2)

Composite Model estimation results:

Init log likelihood:	-6958.425
Final log likelihood:	-4794.544
Likelihood ratio test for the init. model:	4327.761
Rho-square for the init. model:	0.311

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-1.01	0.18	-5.61	2.02e-08	0.194	-5.2	1.95e-07
ASC_SM	-0.237	0.138	-1.72	0.0859	0.142	-1.67	0.0952
B_CAR_COST	-0.00795	0.00108	-7.35	1.95e-13	0.00151	-5.28	1.3e-07
B_CAR_HIGH_INCOME	0.149	0.0553	2.7	0.00702	0.054	2.76	0.0057
B_CAR_LOW_INCOME	0.196	0.095	2.06	0.0394	0.0933	2.1	0.0359
B_CAR_MALE	0.593	0.0704	8.42	0	0.0733	8.09	6.66e-16
B_CAR_TIME	-0.0131	0.000815	-16	0	0.00165	-7.89	3.11e-15
B_FIRST	0.295	0.0785	3.75	0.000175	0.0819	3.6	0.000317
B_GA	0.681	0.191	3.57	0.000359	0.207	3.29	0.000997
B_HE	-0.00648	0.00108	-6.02	1.79e-09	0.00109	-5.95	2.68e-09
B_SENIOR	-1.73	0.122	-14.2	0	0.114	-15.2	0
B_SM_COST	-0.0118	0.000615	-19.2	0	0.00092	-12.8	0
B_SM_HIGH_INCOME	0.186	0.0459	4.04	5.26e-05	0.0454	4.09	4.27e-05
B_SM_LOW_INCOME	-0.324	0.0628	-5.16	2.54e-07	0.0623	-5.2	1.98e-07
B_SM_MALE	0.0869	0.049	1.77	0.0762	0.0501	1.74	0.0826
B_SM_TIME	-0.0109	0.000881	-12.4	0	0.00183	-5.96	2.48e-09
B_TRAIN_COST	-0.0265	0.00127	-20.8	0	0.00207	-12.8	0
B_TRAIN_HIGH_INCOME	-0.335	0.0732	-4.57	4.83e-06	0.0727	-4.61	4.05e-06
B_TRAIN_LOW_INCOME	0.128	0.0757	1.69	0.0906	0.0771	1.66	0.0969
B_TRAIN_MALE	-0.68	0.0643	-10.6	0	0.0658	-10.3	0
B_TRAIN_TIME	-0.00825	0.000855	-9.65	0	0.0012	-6.85	7.28e-12

Table 3: Estimation results from composite Model (MC)

To compare the models, we use the Cox Test. Both model 1 and model 2 are restrictions of the composite model.

To compare model 1 to the composite model, the null hypothesis is:
H_0 : B_car_male = B_sm_male = B_train_male = 0

The likelihood ratio value is X^2 distributed with K=3 degrees of freedom.

LR = -2*(L(M1) - L(MC))
= -2*(-4851.029 + 4794.544) = 112.97 > 7.815 = $X^2_{0.95,3}$

This means we can reject the null hypothesis with a 95% confidence level.

For the second model, M2, we consider M2 a restriction of the composite model, with the null hypothesis:
H_0: B_car_low_income = B_car_high_income = B_train_low_income = B_train_high_income = B_sm_low_income = B_sm_high_income = 0

The likelihood ratio value is X^2 distributed with K=6 degrees of freedom.

LR = -2*(L(M2) - L(MC))
= -2*(-4828.744 + 4794.544) = 68.4 > 12.592 = $X^2_{0.95,6}$

This again results in a rejection of the null hypothesis with a 95% confidence level. This means neither model 1 nor model 2 were found to be a significant improvement over the composite model. More work should be done to develop a model that properly incorporates the income and gender specific socioeconomic variables.

Non-linear travel time

Files:
*SpecTest_SM_base.**
*SpecTest_SM_piecewise_traveltime.**

Description
I hypothesize that travel time effects traveler's preferences differently based on the range of travel time.   For example, the difference between a 30 minute and 60 minute trip may feel much more costly than the difference between a 90 minute trip and 120 minute trip.
In this analysis, I specify a piecewise linear model that divides travel time into two ranges of values: Less than 90 minutes versus greater than or equal to 90 minutes.

This piecewise linear specification builds off the base model (*SpecTest_SM_base)*, and the travel time ranges are used to estimate coefficients for all modes of transit: car, train, and SwissMetro.

The model is specified as follows:

```
V_car = ASC_car + B_car_time1*CAR_TT1 + B_car_time2*CAR_TT2 + B_car_cost*CAR_CO
+ B_senior*SENIOR

V_train = B_train_time1*TRAIN_TT1 + B_train_time2*TRAIN_TT2 +
B_train_cost*TRAIN_COST + B_he*TRAIN_HE + B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time1*SM_TT1 + B_sm_time2*SM_TT2 + B_sm_cost*SM_COST +
B_he*SM_HE + B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

Where

90 > `CAR_TT1, TRAIN_TT1, SM_TT1`
90 <= `CAR_TT2, TRAIN_TT2, SM_TT2`

Results

Piecewise model estimation results:

Init log likelihood:	-6958.425

Final log likelihood: -4872.225

Likelihood ratio test for the init. model: 4172.399

Rho-square for the init. model: 0.3

Rho-square-bar for the init. model: 0.298

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-2.28	0.48	-4.76	1.95e-06	0.507	-4.51	6.61e-06
ASC_SM	-0.623	0.409	-1.52	0.128	0.412	-1.51	0.131
B_CAR_COST	-0.00854	0.00109	-7.87	3.55e-15	0.00148	-5.79	7.02e-09
B_CAR_TIME1	-0.00516	0.00354	-1.46	0.145	0.00369	-1.4	0.161
B_CAR_TIME2	-0.0145	0.000884	-16.4	0	0.00184	-7.87	3.55e-15
B_FIRST	0.304	0.074	4.11	3.9e-05	0.0778	3.91	9.15e-05
B_GA	0.43	0.19	2.27	0.0232	0.202	2.13	0.0328
B_HE	-0.00592	0.00105	-5.64	1.75e-08	0.00106	-5.61	2.05e-08
B_SENIOR	-1.86	0.116	-16	0	0.109	-17	0
B_SM_COST	-0.0118	0.00062	-19	0	0.000938	-12.6	0
B_SM_TIME1	-0.0234	0.00185	-12.7	0	0.00193	-12.1	0
B_SM_TIME2	-0.00486	0.000973	-4.99	5.91e-07	0.00187	-2.6	0.0094
B_TRAIN_COST	-0.0299	0.00128	-23.3	0	0.00211	-14.1	0
B_TRAIN_TIME1	-0.0258	0.00499	-5.18	2.27e-07	0.00509	-5.07	3.9e-07
B_TRAIN_TIME2	-0.00918	0.000959	-9.57	0	0.00104	-8.85	0

Table 4: Piecewise model estimation results.

These results are compared to those of the base model (*SpecTest_SM_base)*, shown in the appendix.
All of the estimated travel time coefficients are negative, which matches our expectations because the more time travel takes, the less attractive it tends to be.
For train and SwissMetro, the estimated travel time coefficients for less than 90 minutes of travel are more negative than than the estimated travel time coefficients for more than 90 minutes of travel. This is not the case for the car, indicating that travelers' preferences are more similar for train and SwissMetro versus car, at least in relation to travel time.

I perform the likelihood ratio test where the restricted model is the one with linear travel time (the base model) and the unrestricted model is the one with piecewise linear travel time.

The null hypothesis is then:
H_0:

```
B_car_time1 = B_car_time2
B_train_time1 = B_train_time2
B_sm_time1 = B_sm_time2
```

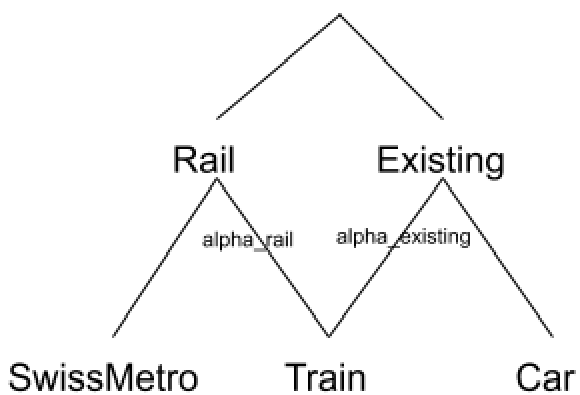
There are K=3 degrees of freedom. The log likelihood ratio test is X^2 distributed and yields

$LR = -2*(-4913.23 + 4872.225) = 82.01 > 7.815 = X^2_{0.95, 3}$

We can therefore reject the null hypothesis of linearly specified travel times at a 95% confidence level. This superior piecewise linear approximation will be used in the further iterations of model development.

Part 2

The best model found for the SwissMetro case due to the analysis in this report is a cross-nested logit model that was estimated with variable alpha values and has the following structure.



The utility expressions are specified as:

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

See below for discussion of this model, as well as the methods and tests used to reach this result. See *MEV_SM_base_CNL_var.** for the model specification code and results.

Estimation of a Nested Logit Model (1)

Files:
*MEV_SM_piecewise_traveltime_NL.**

This model builds on top of the previous section and finding that the piecewise linear approximation for the travel time coefficients resulted in a better model specification than the simpler linear approximation (see *SpecTest_SM_piecewise_traveltime.*)*.
This model is used as the base and restricted logit model in the tests of this nested logit model.

I hypothesize that the train and SwissMetro transit alternatives are quite similar versus the car alternative because they are both rail based, do not require the traveler to do the work of driving, and involve interaction with the public. Meanwhile the car alternative involves driving in one's own private car and gives the traveler more control over when to leave, when to stop, who to interact with, and the trip overall. I hypothesize that the train and SwissMetro are closer substitutes to each other and share unobserved variables and therefore correlated sources of error in the model. For these reasons, train and SwissMetro are nested together in a nested logit model, with car travel alone in its own nest.

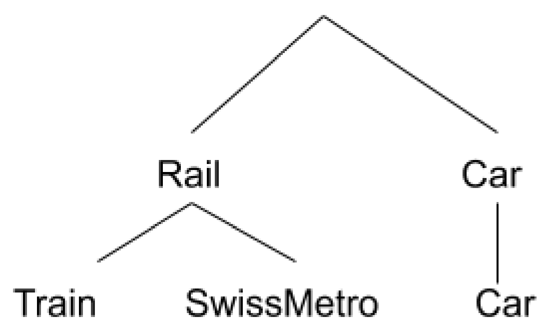


Figure: Correlation structure of NL model

The expressions for the systematic utility functions for each of the travel alternatives use the previously developed piecewise linear approximation for travel time (*SpecTest_SM_piecewise_traveltime.*)* and are defined as follows:

```
V_car = ASC_car + B_car_time1*CAR_TT1 + B_car_time2*CAR_TT2 + B_car_cost*CAR_CO
+ B_senior*SENIOR

V_train = B_train_time1*TRAIN_TT1 + B_train_time2*TRAIN_TT2 +
B_train_cost*TRAIN_COST + B_he*TRAIN_HE + B_ga*GA + B_firstclass*FIRST
```

```
V_sm = ASC_sm + B_sm_time1*SM_TT1 + B_sm_time2*SM_TT2 + B_sm_cost*SM_COST +
B_he*SM_HE + B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

Where

90 > `CAR_TT1, TRAIN_TT1, SM_TT1`
90 <= `CAR_TT2, TRAIN_TT2, SM_TT2`

The scale parameter for the upper level and car nest is fixed to 1 so that the scale parameter for the rail nest can be estimated.

Results

Init log likelihood:	-6958.425
Final log likelihood:	-4858.097
Likelihood ratio test for the init. model:	4200.656
Rho-square for the init. model:	0.302

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-1.71	0.292	-5.84	5.1e-09	0.335	-5.1	3.45e-07
ASC_SM	-0.0232	0.114	-0.203	0.839	0.117	-0.199	0.842
B_CAR_COST	-0.00685	0.00108	-6.34	2.29e-10	0.00148	-4.62	3.89e-06
B_CAR_TIME1	0.00465	0.00339	1.37	0.171	0.00371	1.25	0.211
B_CAR_TIME2	-0.0131	0.000881	-14.9	0	0.00186	-7.06	1.69e-12
B_FIRST	0.292	0.0727	4.01	6.06e-05	0.0783	3.73	0.000195
B_GA	0.513	0.188	2.73	0.00633	0.2	2.56	0.0104
B_HE	-0.00164	0.00035	-4.68	2.82e-06	0.000456	-3.59	0.000327
B_SENIOR	-0.518	0.0688	-7.53	5.04e-14	0.114	-4.54	5.53e-06
B_SM_COST	-0.0128	0.000585	-22	0	0.000927	-13.9	0
B_SM_TIME1	-0.00729	0.00116	-6.3	2.91e-10	0.00189	-3.85	0.000117
B_SM_TIME2	-0.00345	0.000548	-6.29	3.12e-10	0.000869	-3.97	7.19e-05
B_TRAIN_COST	-0.0202	0.00102	-19.8	0	0.00181	-11.2	0
B_TRAIN_TIME1	-0.00569	0.00171	-3.33	0.000877	0.00225	-2.53	0.0114
B_TRAIN_TIME2	-0.00321	0.000498	-6.45	1.13e-10	0.000824	-3.9	9.62e-05
MU_rail	3.79	0.446	8.5	0	0.791	4.8	1.62e-06

Table 5: Estimation results for nested logit model.

Note that the model was estimated multiple times with different starting values for MU_rail in order to decrease the chance of only finding local optima.

To evaluate this model versus the restricted logit model (*SpecTest_SM_piecewise_traveltime.*)*, we test the null hypothesis that the scale parameter is 1, which makes this model equivalent to the restricted logit mode.
H_0: MU_rail = 1

Since there is only one restriction, we can use a t-test, computed as follows.

t-statistic = (MU_rail - 1)/std. error = (3.79 - 1)/0.791 = 3.53.

The t-statistic with respect to 1 is 3.53 which indicates that MU_rail is significantly different from 1 and that therefore there is significant correlation between the rail options of train and SwissMetro.

Overall, the estimated parameters are sensible and match intuition.
The MU_rail value was expected to be more than 1 because it is in a lower level nest, and the upper level nest scale value (MU) was fixed to 1, and we expect MU/MU_rail < 1.
All of the estimated cost coefficients are negative, which makes sense because more expensive options are less appealing.
The estimated coefficient for headway is negative as well, which also makes sense because greater headway is an inconvenience that adds to travel time.
The positive values for the coefficients for the first class and GA ticket dummy variables match expectations because they make train and SwissMetro alternatives less expensive and more convenient.
The small negative values for the SwissMetro and Car alternative specific constants reflect a slight preference for the train alternative, with all other factors excluded.

Due to the success of this model, I wanted to extend it to a cross-nested model. Unfortunately, this was not possible due to limitations with the biogeme modeling software, as it resulted in too complex a model for estimation. The model was still written (see *MEV_SM_piecewise_traveltime_CNL_fix.ipynb*).

In order to continue towards CNL development, I returned to the base model (*SpecTest_SM_base*) that was previously developed and built on in Part 1.

Estimation of a Nested Logit Model (2)

Files:
*MEV_SM_base_NL.**

For the development of this model, the hypotheses regarding the relationship between the transit alternatives and their correlation structure are the same as described above.

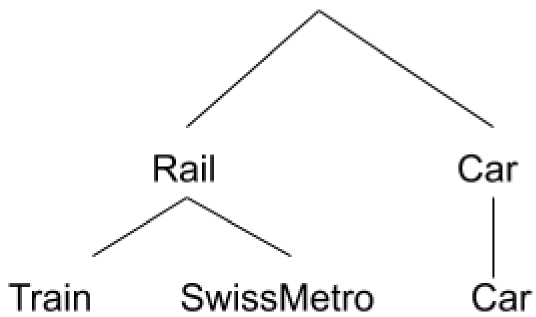


Figure: Correlation structure of NL model

The utility expressions are specified as follows (they were tested and validated in case study 3):

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

Results

Init log likelihood:	-6958.425
Final log likelihood:	-4884.805
Likelihood ratio test for the init. model:	4147.238
Rho-square for the init. model:	0.298

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.252	0.0926	-2.72	0.00651	0.102	-2.48	0.0131
ASC_SM	0.0239	0.0279	0.855	0.393	0.0277	0.863	0.388
B_CAR_COST	-0.00601	0.00106	-5.68	1.37e-08	0.00149	-4.04	5.36e-05
B_CAR_TIME	-0.0111	0.000793	-13.9	0	0.00159	-6.97	3.14e-12
B_FIRST	0.244	0.0722	3.38	0.000713	0.0785	3.11	0.00185
B_GA	0.64	0.187	3.42	0.000635	0.2	3.2	0.00137
B_HE	-0.00156	0.000321	-4.87	1.14e-06	0.000404	-3.87	0.000109
B_SENIOR	-0.49	0.0613	-7.98	1.33e-15	0.0977	-5.01	5.34e-07
B_SM_COST	-0.0125	0.000582	-21.4	0	0.000925	-13.5	0
B_SM_TIME	-0.00401	0.000566	-7.09	1.37e-12	0.00106	-3.78	0.000155
B_TRAIN_COST	-0.0194	0.000982	-19.8	0	0.0017	-11.4	0
B_TRAIN_TIME	-0.00287	0.000445	-6.44	1.23e-10	0.000763	-3.76	0.000171
MU_rail	4.04	0.442	9.15	0	0.76	5.32	1.04e-07

The estimated coefficients again match expectations for the same reasons as *MEV_SM_piecewise_traveltime_NL.**. The cost, time, and headway coefficients are negative while the coefficients for the GA and first class dummy variables are positive.

In order to test whether the scale parameter is statistically significant and whether the hypothesized nesting structure is therefore justified, the t-statistic is again used. The null hypothesis is again:
H_0: MU_rail = 1

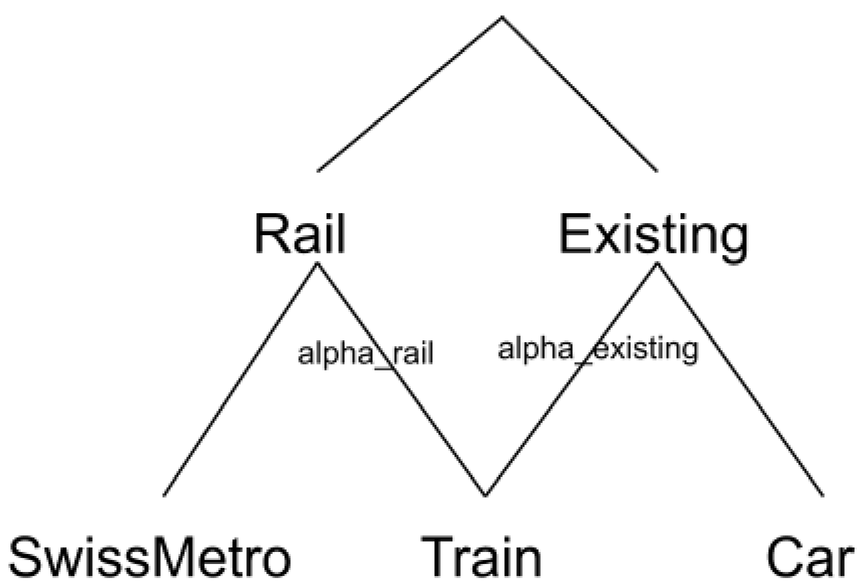
t-statistic = (MU_rail - 1)/std. error = (4.04 - 1)/0.76 = 4.

Since MU_rail is statistically different than 1, we reject the null hypothesis and continue to assume and use the nested correlation structure for this model.

Estimation of a Cross-Nested Logit Model with fixed alpha values

Files:
*MEV_SM_base_NL.**
*MEV_SM_base_CNL_fix.**

This cross-nested logit model builds on the work for the simpler nested logit specification (*MEV_SM_base_NL.**).



In this specification, there are two nests:

1. “Rail”: Rail based travel options which are SwissMetro and Train.
2. “Existing”: Existing travel alternatives of Train and Car.

In this new cross-nested model, the Train alternative is in a nest shared with SwissMetro (“Rail”) and Car (“Existing”).

The similarities and correlation between the “Rail” options of SwissMetro and Train were discussed and numerically validated previously.

I hypothesize that the choice for the Train travel alternative is also related to the Car alternative and that they share unobserved variables. This is because both Train and Car are existing travel options, and travelers may feel more comfortable with continuing to use the travel alternatives they already know and are used to rather than trying out new options such as SwissMetro. In other words, there may be unobserved cost variables for traveler’s changing travel modes from what they already know and feel comfortable with.

For this model, the utility expressions are the same as for *MEV_SM_base_NL*.

There are now two lower nests, and the scale parameters for these two lower nests are not necessarily the same. To allow them to take different values, the top scale parameter is again fixed to 1 (MU=1) and the lower scale parameters (MU_rail and MU_existing) are estimated variables. These scale parameters are both expected to have values greater than 1.

For this model, the alpha values (alpha_rail and alpha_existing) that represent the weight of the train alternative’s membership in the rail versus existing lower logit model nests are fixed. I hypothesize that the choice to take the train is more similar and correlated with the choice to take SwissMetro rather than the choice to take a car, so I fix the alphas to represent this:

```
alpha_rail = 0.6
alpha_exisitng =  0.4
```

In the estimation, multiple simulations are run with different starting values for the scale parameters in order to avoid results from local optima that do not reflect the global optimum. Different optima were not found.

Results

Init log likelihood:	-6958.425
Final log likelihood:	-4892.921
Likelihood ratio test for the init. model:	4131.007
Rho-square for the init. model:	0.297

Rho-square-bar for the init. model: 0.295

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.586	0.117	-4.99	6.07e-07	0.127	-4.63	3.66e-06
ASC_SM	-0.267	0.078	-3.42	0.000623	0.0779	-3.42	0.00062
B_CAR_COST	-0.0071	0.00104	-6.85	7.16e-12	0.00153	-4.63	3.63e-06
B_CAR_TIME	-0.012	0.000859	-13.9	0	0.00153	-7.81	6e-15
B_FIRST	0.248	0.0713	3.48	0.000508	0.0765	3.24	0.0012
B_GA	0.491	0.176	2.79	0.00519	0.183	2.68	0.0074
B_HE	-0.00409	0.000724	-5.66	1.52e-08	0.000755	-5.42	5.85e-08
B_SENIOR	-1.05	0.0929	-11.3	0	0.0975	-10.7	0
B_SM_COST	-0.0114	0.000676	-16.9	0	0.000985	-11.6	0
B_SM_TIME	-0.00913	0.0007	-13	0	0.00128	-7.14	9.59e-13
B_TRAIN_COST	-0.0249	0.00159	-15.6	0	0.00238	-10.5	0
B_TRAIN_TIME	-0.00703	0.000656	-10.7	0	0.00103	-6.83	8.51e-12
MU_existing	1.12	0.0725	15.4	0	0.106	10.5	0
MU_rail	2.9	0.378	7.66	1.89e-14	0.346	8.39	0

All of the estimated coefficient values are sensible match expectations. The only coefficient that changed sign from the previous nested version of this model (*MEV_SM_base_NL.*)* is an alternative specific constant.

The scale parameters for the two lower logit nests (MU_rail and MU_existing) took on very different values. It is interesting to see that the estimated scale for the rail nest is more than twice the scale value for the existing nest. A higher scale means lower variance, which helps validate the hypothesis that the choice for the train alternative is more closely correlated to the other rail option rather than the other existing option.

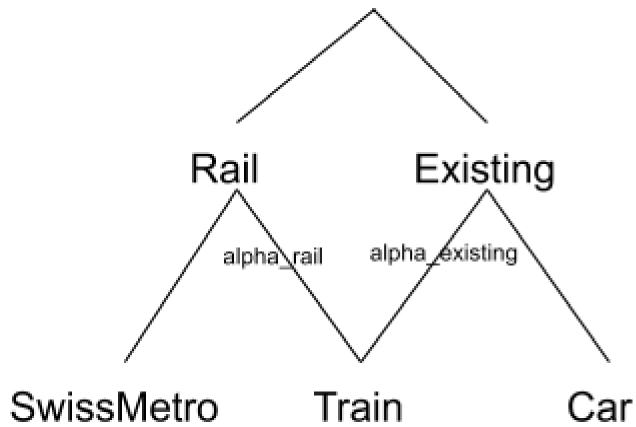
The previously estimated nested model (*MEV_SM_base_NL.*)* can be considered a restriction of this CNL, where the train alternative can only belong to the rail nest. We would then test whether the CNL is an improvement over this simpler nested model. The test statistic to use to compare the models would be the log likelihood ratio. However, we can immediately see that the log likelihood for this CNL (-4892.921) is worse than for its restriction (-4884.805), and reject this CNL.

Estimation of a Cross-Nested Logit Model with variable alpha values

Files:
*MEV_SM_base_NL.**
*MEV_SM_base_CNL_var.**

This model assumes the same hypotheses as the previous, but allows the alpha values to take on variable values in order to estimate values that are more successful than the previous assumptions of fixed values.

The cross-nested structure is the same as before,



as are the utility expressions:

```
V_car = ASC_car + B_car_time*CAR_TT + B_car_cost*CAR_CO + B_senior*SENIOR

V_train = B_train_time*TRAIN_TT + B_train_cost*TRAIN_COST + B_he*TRAIN_HE +
B_ga*GA + B_firstclass*FIRST

V_sm = ASC_sm + B_sm_time*SM_TT + B_sm_cost*SM_COST + B_he*SM_HE +
B_senior*SENIOR + B_ga*GA + B_firstclass*FIRST
```

The alpha values are variable.

Multiple simulations are run, with the scale variables starting at different values, in order to avoid results with local optima that are not globally optimal.

Results

Init log likelihood:	-6530.284
Final log likelihood:	-4867.592
Likelihood ratio test for the init. model:	3325.383
Rho-square for the init. model:	0.255

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.346	0.0946	-3.66	0.00025	0.107	-3.25	0.00117
ASC_SM	-0.11	0.0318	-3.47	0.000512	0.0417	-2.64	0.00817
B_CAR_COST	-0.00581	0.00103	-5.66	1.5e-08	0.00152	-3.83	0.000131
B_CAR_TIME	-0.0102	0.000846	-12.1	0	0.0015	-6.79	1.11e-11
B_FIRST	0.235	0.0702	3.35	0.000802	0.0754	3.12	0.0018
B_GA	0.603	0.179	3.37	0.00074	0.196	3.08	0.00208
B_HE	-0.00127	0.00027	-4.71	2.46e-06	0.000366	-3.49	0.000491
B_SENIOR	-0.344	0.0449	-7.66	1.82e-14	0.072	-4.78	1.77e-06
B_SM_COST	-0.0119	0.000631	-18.9	0	0.000969	-12.3	0
B_SM_TIME	-0.00322	0.000471	-6.83	8.3e-12	0.000847	-3.8	0.000143
B_TRAIN_COST	-0.019	0.00109	-17.5	0	0.00175	-10.9	0
B_TRAIN_TIME	-0.00226	0.000401	-5.64	1.72e-08	0.000697	-3.24	0.0012
MU_existing	1.15	0.0991	11.6	0	0.168	6.86	7.07e-12
MU_rail	7.12	0.923	7.71	1.29e-14	1.44	4.93	8.03e-07
alpha_rail_train	0.892	0.0198	45.1	0	0.029	30.8	0

The difference in the estimated coefficients for the scale parameters (for the “rail” versus “existing” logit nests) is more pronounced than for the previous CNL with fixed alphas. The alpha value for rail is greater than the previously assumed fixed value of 0.6. This matches the hypothesis that the train alternative’s membership in the rail logit nest is more important than its membership in the logit nest for existing transit.

The other estimated coefficients match expectations. All cost, time, and headway coefficients are negative, and the coefficients for the GA and first class dummy variables are positive.

The previously estimated simple nested model (*MEV_SM_base_NL.**) is a restriction of this CNL. This restriction is equivalent to setting MU_existing to 1, alpha_rail to 1, and alpha_existing to 0. To validate that the CNL is a better model than this simpler nested restriction, we test the null hypothesis that assumes the restriction.

H_0: MU_existing = 1; alpha_existing = 0

The test statistic used is the log likelihood ratio, which is X^2 distributed with 2 degrees of freedom.

$LR = -2(L_{restricted} - L_{unrestricted}) = -2*(-4884.805 + 4867.592) = 34.426 > 5.991 = X^2_{0.95,\ 2}$

We can reject the null hypothesis with a 95% confidence level and accept this CNL as an improved model.

In order to test whether this CNL is an improvement over the nested logit model that was specified with piecewise linear travel times (*SpecTest_SM_piecewise_traveltime)*, I would create a composite model of the two, and this composite model would necessarily be a cross-nested logit model. However, estimation of this model would experience the same limitations in the biogeme software as before, so these two models cannot be properly compared. So I accept this CNL with variable alphas and linear utility function specifications as the “best” model.

Code

The code used in this report is submitted online in a zip file.

Appendix

The estimation results for the base model used in part 1:

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	-0.216	0.149	-1.45	0.148	0.161	-1.34	0.18
ASC_SM	0.123	0.119	1.03	0.301	0.12	1.02	0.306
B_CAR_COST	-0.00751	0.00107	-7.01	2.39e-12	0.00151	-4.99	6.11e-07
B_CAR_TIME	-0.0131	0.000812	-16.1	0	0.00164	-7.95	1.78e-15
B_FIRST	0.262	0.0735	3.56	0.000375	0.0789	3.31	0.000925
B_GA	0.518	0.188	2.75	0.00588	0.194	2.67	0.00764
B_HE	-0.00589	0.00105	-5.63	1.83e-08	0.00105	-5.61	2.01e-08
B_SENIOR	-1.86	0.116	-16.1	0	0.107	-17.4	0
B_SM_COST	-0.0116	0.000617	-18.8	0	0.000931	-12.5	0
B_SM_TIME	-0.011	0.000877	-12.5	0	0.00181	-6.06	1.36e-09
B_TRAIN_COST	-0.0293	0.00128	-23	0	0.0021	-14	0
B_TRAIN_TIME	-0.00881	0.000855	-10.3	0	0.0012	-7.35	2.03e-13

Appendix Table 1.