Part 1: Specification Testing for Multinomial Logit Models

1 Objectives of Part 1 of the Case Study

The topic of Part 1 of this case study is to test different hypotheses regarding model specifications. The objectives can be summarized as follows:

- Illustration of the market segmentation concept and related testing.
- Testing of non-nested hypotheses using the Cox test.
- Testing of non-linear specifications using the piecewise linear approximation, the power series expansion and the Box-Cox transformation methods.

2 Theoretical Reminder

The statistical tests which will be used in this case study are t -tests and likelihood ratio tests. We give in the following a short reminder of these concepts, leaving the specific details to the examples used in the case study.

2.1 Market segmentation

The *market segmentation* procedure captures systematic variations of the coefficients in different subgroups of the sample. Let's assume that our sample can be segmented in groups $g = 1, ..., G$ with respect to a certain variable (e.g., age or income). We denote by N_g the sample size of market segment g and hence

$$\sum_g N_g = N$$

where N is the full sample size. The null hypothesis, corresponding to the case of no variation among the market segments, is:

$$H_0 : \beta^1 = ... = \beta^g = ... = \beta^G$$

The statistic used to test the null hypothesis is based on the likelihood ratio test and is given by:

$$-2(\mathcal{L}_N(\hat{\beta}) - \sum_g \mathcal{L}_{N_g}(\hat{\beta}^g))$$

where the first term represents the log likelihood for the restricted model using the same parameter set on the full sample and the terms in the summation represent the log likelihood values of the unrestricted models estimated on the sample segments. This statistic is χ^2 distributed with the degrees of freedom given by:

$$df = \sum_g K_g - K$$

where K_g is the number of coefficients in segment g and K is the number of coefficients in the restricted model.

2.2 Test of Non-Nested Hypothesis

Several tests for discrete choice models have the characteristic of being based on the so called *nested* hypothesis. The idea actually is to compare two different models (restricted and unrestricted) where the first one represents the null hypothesis and is obtained as a special case of the unrestricted one. There are situations where the models we aim to compare are not nested, meaning that we cannot obtain one as a restricted version of the other. In these cases, the standard likelihood ratio test cannot be applied. We describe in the following three different procedures to compare such models, which actually can be interpreted as different *extensions* of a common base model M_b. The interested reader can find more details on this topic in Ben-Akiva and Lerman (1985).

Composite model test

In this procedure, the goal is to specify a composite model (we call it M_C). The number of parameters in such a model is larger than the two original extensions, called M_1 and M_2. The idea is that if both models M_1 and M_2 are derived from M_C as two special cases, then we can perform two likelihood ratio tests for each of the two restricted versions against the composite model.

The $\bar{\rho}^2$ and J tests

In this procedure, we use the adjusted likelihood ratio index $\bar{\rho}^2$ as a goodness-of-fit measure to test non-nested hypothesis on the two alternative specifications M_1 and M_2. The first step consists of computing the index for each of the two models as follows:

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}(\hat{\beta}) - K}{\mathcal{L}(0)}$$

where $\mathcal{L}(\hat{\beta})$ is the log likelihood value at the estimated coefficients, K is the number of parameters in the model and $\mathcal{L}(0)$ is the null log likelihood (case of all parameters being equal to zero). Now, under the assumption that model M_1 is the true specification, we make use of the following asymptotic inequality:

$$\Pr(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq \Phi\{-[-2z\mathcal{L}(0) + (K_2 - K_1)]^{\frac{1}{2}}\}$$

where K_1 and K_2 are the number of parameters in the two models, $z > 0$ and Φ the standard normal cumulative distribution function. Practically, such inequality shows that the probability that another model (M_2) has a better goodness-of-fit against the true model M_1 is bounded. In other words, choosing the model with the higher likelihood ratio index bounds the probability of erroneously choosing the incorrect model over the true specification. If we have all the J alternatives for each of the N observations, the above inequality is reduced to:

$$\Pr(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq \Phi\{-[2Nz\ln(J) + (K_2 - K_1)]^{\frac{1}{2}}\}$$

2.3 Tests of Non-Linear Specifications

Usually discrete choice models are specified with linear-in-parameters utility functions. This assumption arises from the fact that often we do not have any *a priori* knowledge justifying the non-linearity assumption. However, such a specification gives us a certain degree of flexibility allowing for non-linearities in the variables. Different methodologies have been developed to test such non-linear (in variables) specifications. We give in the following a description and example of three of these methods: the piecewise linear approximation, the power series method and the Box-Cox transformation.

Piecewise linear approximation

The idea in this approach is to split the range of values for a certain variable into n parts and associate a different coefficient with each of them. We actually want to test if the coefficient for that variable assumes different values for different values (or ranges of values) of the variable itself. In order to perform such a test, a χ^2 test statistic is used where the null hypothesis consists of the assumption that all the coefficients for all the splitted variables are equal to each other. Problems with this method arise from the loss of degrees of freedom as a consequence of a large number of variable's ranges. Moreover, it is possible to have only few observations for some of the ranges. Finally, the range endpoints have to be decided arbitrarily.

The power series expansion

Here the non-linearity is expressed by a power series expansion which includes the linear specification as a special case. Normally, only a few terms of the expansion are considered, as they are highly correlated among themselves. Both the power series and the piecewise linear approximation methods involve estimations of models that are linear in the unknown parameters.

The Box-Cox transformation

With this method, it is possible to test for non-linearities of variables that are non-linear in the unknown parameters. The Box-Cox transformation applies to non-negative values and is given by the following expression:

$$\frac{x^\lambda - 1}{\lambda}, \text{with } x, \lambda \geq 0$$

where λ is an unknown parameter. This transformation includes the linear case when $\lambda = 1$. The objective is to find the value of the λ parameter that maximizes the likelihood ratio index.

3 Datasets

For this case study, you can choose between the *Residential Telephone Services*, and the *Swissmetro* datasets. A detailed description of each dataset can be found in Appendix A. Examples of model specifications are provided for both datasets in the following sections.

The Swissmetro dataset

This dataset consists of survey data collected on the train between St. Gallen and Geneva, Switzerland, during March 1998. The respondents provided information in order to analyze the impact of the modal innovation in transportation, represented by the Swissmetro, a revolutionary mag-lev underground system, against the usual transport modes represented by car and train.

The residential telephone services dataset

A household survey was conducted in 1984 for a telephone company among 434 households in Pennsylvania. The dataset involves choices among five calling plans and consists of various alternative-specific and socio-economic variables. It was originally used to develop a model system to predict residential telephone demand (Train and Ben-Akiva (1987)).

4 Your tasks

The procedure you should follow to work on the case study can be summarized as follows:

- select a dataset of interest, as you will be working on only one of two datasets;
- formulate hypothesis: selection of the explanatory variables and how they affect the utilities (generic and/or alternative-specific), inclusion of socio-economic variables, etc.;

- perform the estimation of the related model. Give an interpretation of the obtained results and check if they are plausible;
- try to modify the model, formulating different kinds of hypotheses (nested/non- nested) and specifications (linear/non-linear);
- apply different test procedures based on the alternative hypotheses / specifications.
- Give an interpretation of the obtained results remembering that model specification is an iterative process.

5 Report Content

Your final report should contain:

- the presentation of THREE different specifications
  - select three among different options of market segmentation, non-nested hypothesis and nonlinear specifications: piecewise-linear, box-cox, power series
- discussion of the relative motivation and reasoning behind these specifications;
- the results of some tests used to arrive at your presented specification;
- a discussion of the estimated parameter values;
- the final BIOGEME outputs for all discussed models.

Choice of Residential Telephone Services Case

Market Segmentation

Files to use with Biogeme:
Model files: *SpecTest_Tel_low_inc.py*, *SpecTest_Tel_med_inc.py*, *SpecTest_Tel_high_inc.py*, *MNL_Tel_socioec.py*
Data file: *telephone.dat*

We test if there is a taste variation across market segments. We define different segments based on income and divide the population into three income groups. We estimate separate models for each income group using the same model specification, namely *MNL_Tel_socioec.py* used in the logit case study, and compare the estimation results with a model based on the complete dataset. The results in terms of final log-likelihood are summarized in Table 1. Note that for the low income group, the model has one parameter less (6 instead of 7) because $\mathsf{ASC_{EF}}$ has been fixed to zero. This is due to the fact that no subject with low income chose that alternative, so $\mathsf{ASC_{EF}}$ is impossible to estimate.

The null hypothesis is of no taste variation across the market segments, that is

$$H_0 \ : \ \beta_{HI} = \beta_{MI} = \beta_{LI}.$$

Performing a likelihood ratio test,

$$\begin{aligned} LR &= -2(L_N(\hat{\beta}) - \sum_{g=1}^{G} L_{N_g}(\hat{\beta}^g)) \\ &= -2(-468.791 + 120.103 + 297.990 + 46.668) = 8.060 \\ \chi^2_{0.95,13} &= 22.360 \end{aligned}$$

We can conclude that the null hypothesis cannot be rejected, that is, market segmentation on income does not exist.

Model	Definition	Log-likelihood	Nb. of Coefficients
Low Income	$\mathrm{Income} < 10000$	-120.103	6
Medium Income	$10000 < \mathrm{Income} < 40000$	-297.990	7
High Income	$\mathrm{Income} > 40000$	-46.668	7
Pooled Data Restricted Model	All	-468.791	7

Table 1: Results for the market segmentation test

Test of Non-Nested Hypotheses

In discrete choice analysis, we often perform tests based on the so-called nested hypotheses, which means that we specify two models such that the first one (the restricted model) is a special case of the second one (the unrestricted model). For this type of comparison, the classical likelihood ratio test can be applied. However, there are situations in which we aim at comparing models which are not nested, meaning that one model cannot be obtained as a restricted version of the other. One way to compare two non-nested models is to build a composite model from which both models can be derived. We can thus perform two likelihood ratio tests for each of the restricted models against the composite model. This procedure is known as the Cox test of separate families of hypothesis.

Cox Test

Files to use with Biogeme:
Model file: *SpecTest_Tel_M1.py*, *SpecTest_Tel_M2.py*
SpecTest_Tel_MC.py
Data file: *telephone.dat*

The Cox test is described in detail in Ben-Akiva and Lerman (1985), pages 171-174, and in the Textbook of the course, in section "Tests of Non-Nested Hypothesis". Assume that we want to test a model M_1 against another model M_2 (and one model is not a restricted version of the other). We start by generating a composite model M_C such that both models M_1 and M_2 are restricted cases of M_C. We then test M_1 against M_C and M_2 against M_C using the likelihood ratio test. There are three possible outcomes of this test:

- One of the two models is rejected. Then we keep the one that is not rejected.
- Both models are rejected. Then better models should be developed.
- Both models are accepted. Then we choose the model with the higher $\bar{\rho}^2$ index.

The deterministic parts of the utility functions for each of the three model specifications are:

1. M_1

$$
\begin{aligned}
V_{BM} &= ASC_{BM} + \beta_{Mcost} cost_{BM} \\
V_{SM} &= \beta_{Mcost} cost_{SM} \\
V_{LF} &= ASC_{LF} + \beta_{Fcost} cost_{LF} \\
V_{EF} &= ASC_{EF} + \beta_{Fcost} cost_{EF} \\
V_{MF} &= ASC_{MF} + \beta_{Fcost} cost_{MF}
\end{aligned}
$$

2. M_2

$$
\begin{aligned}
V_{BM} &= ASC_{BM} + \beta_{cost} cost_{BM} \\
V_{SM} &= \beta_{cost} cost_{SM} \\
V_{LF} &= ASC_{LF} + \beta_{cost} cost_{LF} + \beta_{users} users \\
V_{EF} &= ASC_{EF} + \beta_{cost} cost_{EF} + \beta_{users} users \\
V_{MF} &= ASC_{MF} + \beta_{cost} cost_{MF} + \beta_{users} users
\end{aligned}
$$

3. M_c

$$
\begin{aligned}
V_{BM} &= ASC_{BM} + \beta_{Mcost} cost_{BM} \\
V_{SM} &= \beta_{Mcost} cost_{SM} \\
V_{LF} &= ASC_{LF} + \beta_{Fcost} cost_{LF} + \beta_{users} users \\
V_{EF} &= ASC_{EF} + \beta_{Fcost} cost_{EF} + \beta_{users} users \\
V_{MF} &= ASC_{MF} + \beta_{Fcost} cost_{MF} + \beta_{users} users
\end{aligned}
$$

Model	Nb. of parameters	Log-likelihood	$\bar{\rho}^2$
M_1	6	-476.040	0.140
M_2	6	-471.151	0.148
M_C	7	-467.804	0.153

Table 2: Results from the non-nested hypothesis test

The estimation results of the different models are summarized in Table 2. We first compare the M_1 model specification against the composite model M_C by means of a likelihood ratio test:

$$
\begin{gathered}
H_0 : \beta_{\text{users}} = 0 \\
-2(L(\hat{\beta}_{M1}) - L(\hat{\beta}_{MC})) = -2(-476.040 + 467.804) = 16.472 \\
\chi^2_{0.95,1} = 3.841 < 16.472
\end{gathered}
$$

We can therefore reject the null hypothesis of not including socio-economic variables. We then compare M_2 against M_C:

$$
\begin{gathered}
H_0 : \beta_{Mcost} = \beta_{Fcost} \\
-2(L(\hat{\beta}_{M2}) - L(\hat{\beta}_{MC})) = -2(-471.151 + 467.804) = 6.694 \\
\chi^2_{0.95,1} = 3.841 < 6.694
\end{gathered}
$$

We can therefore reject the null hypothesis of generic coefficients. Since both models are rejected, we need to develop better models. Had both models been accepted, we could have used $\bar{\rho}^2$ to choose which model to keep.

The adjusted likelihood ratio index $\bar{\rho}^2$ is computed as follows (it is provided in the Biogeme result file):

$$
\bar{\rho}^2 = 1 - \frac{L(\hat{\beta}) - K}{L(0)}
$$

So, for the two models M_1 and M_2, we obtain respectively:

$$
\begin{aligned}
\bar{\rho_1}^2 &= 0.140 \\
\bar{\rho_2}^2 &= 0.148
\end{aligned}
$$

Tests of Non-Linear Specifications

In the previous case study, the models were specified with linear in parameter formulations of the deterministic parts of the utilities (parameters that remain constant throughout the whole range of the values of each variable). However, in some cases, non-linear specifications may be more justified (e.g. sensitivity to cost may not be the same in all cost ranges). In this section, we test three different non-linear specifications of the deterministic utility functions (see Ben-Akiva and Lerman, 1985, pages 174-179). Namely, piecewise linear approximation, power series method and Box-Cox transformation are used below. We have used the logit model with alternative specific cost coefficients as the base model (*SpecTest_Tel_M1.py*).

Piecewise Linear Approximation

Files to use with Biogeme:
Model file: *SpecTest_Tel_piecewise.py*
Data file: *telephone.dat*

In the first model, we assume that the coefficient of measured cost assumes different values for different ranges of the cost variable. The full range of values for the measured cost variable is \$3.28 to \$433.5. We split the range of values for cost_i (which is $\mathsf{cost}_i \in [3.28, 433.5]$, expressed in dollars) into three different intervals: $\mathsf{cost}_{i1} \in [0, 10]$, $\mathsf{cost}_{i2} \in (10, 50]$ and $\mathsf{cost}_{i3} > 50$. The selection of these ranges is based on a priori hypotheses of the user behavior and distribution of cost in the observed sample. The reader is encouraged to experiment with different ranges. An extract from the Biogeme model file to code the ranges of costs is presented in Figure 1.

The deterministic utility functions are

$$
\begin{aligned}
V_{BM} &= ASC_{BM} + \beta_{Mcost1}\mathsf{cost}_{BM1} + \beta_{Mcost2}\mathsf{cost}_{BM2} + \beta_{Mcost3}\mathsf{cost}_{BM3} \\
V_{SM} &= \beta_{Mcost1}\mathsf{cost}_{SM1} + \beta_{Mcost2}\mathsf{cost}_{SM2} + \beta_{Mcost3}\mathsf{cost}_{SM3} \\
V_{LF} &= ASC_{LF} + \beta_{Fcost}\mathsf{cost}_{LF} \\
V_{EF} &= ASC_{EF} + \beta_{Fcost}\mathsf{cost}_{EF} \\
V_{MF} &= ASC_{MF} + \beta_{Fcost}\mathsf{cost}_{MF}
\end{aligned}
$$

```
    // Define here arithmetic expressions for name
    // that are not directly available from the data
one  = DefineVariable('one',1)
cost11  = DefineVariable('cost11', min(cost1, 10))
cost12  = DefineVariable('cost12', max(0,min(cost1 - 10 ,40)))
cost13  = DefineVariable('cost13', max(0,cost1 - 50))
cost21  = DefineVariable('cost21', min(cost2 ,10))
cost22  = DefineVariable('cost22', max(0,min(cost2 - 10 ,40)))
cost23  = DefineVariable('cost23', max(0,cost2 - 50))
```

Figure 1: Biogeme snapshot for the piecewise linear approximation

Piecewise linear approximation

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	$\mathsf{ASC}_{\mathrm{BM}}$	-0.613	0.152	-4.03
2	$\mathsf{ASC}_{\mathrm{LF}}$	-0.631	0.500	-1.26
3	$\mathsf{ASC}_{\mathrm{EF}}$	-0.843	0.869	-0.97
4	$\mathsf{ASC}_{\mathrm{MF}}$	-0.261	0.640	-0.41
5	β_{Mcost1}	-0.294	0.0661	-4.44
6	β_{Mcost2}	-0.149	0.0665	-2.23
7	β_{Mcost3}	-1.38	0.644	-2.14
8	β_{Fcost}	-0.105	0.0217	-4.84

Summary statistics
Number of observations = 434
$\mathcal{L}(0) = -560.250$
$\mathcal{L}(\hat{\beta}) = -474.703$
$\bar{\rho}^2 = 0.138$

Table 3: Estimation results for the piecewise linear approximation

The results shown in Table 3 indicate that the sensitivity to measured cost becomes less important in the range $10 < \mathsf{cost}_i < 50$ compared to the range $\mathsf{cost}_i < 10$, but has a steep increase for higher costs. This model has a better goodness-of-fit than the model with linear coefficients in general. To test whether or not the improvement in goodness-of-fit is statistically significant, we need to perform a likelihood ratio test between the two different specifications.

The null hypothesis in this case is

$$H_0 : \beta_{\text{Mcost1}} = \beta_{\text{Mcost2}} = \beta_{\text{Mcost3}}$$

The χ^2 statistic for this null hypothesis is as follows:

$$-2(L(\hat{\beta}_R) - L(\hat{\beta}_U)) = -2(-476.040 + 474.703) = 2.674$$
$$\chi^2_{0.95,2} = 5.991 > 2.674$$

where the restricted model (R) is represented by the linear specification while the unrestricted model (U) corresponds to the piecewise linear specification. The improvement in goodness-of-fit due to the introduction of the piecewise linear specification is not significant and the null hypothesis that the cost coefficient is linear cannot be rejected.

The Power Series Expansion

Files to use with Biogeme:
Model file: *SpecTest_Tel_powerseries.py*
Data file: *telephone.dat*

In this test, we relax the hypothesis of linear coefficients for measured options by assuming a second order power series (a squared term and a linear term).

The corresponding systematic utility functions are

$$
\begin{aligned}
V_{BM} &= ASC_{BM} + \beta_{Mcost1} cost_{BM} + \beta_{Mcost2} cost_{BM}^2 \\
V_{SM} &= \beta_{Mcost1} cost_{SM} + \beta_{Mcost2} cost_{SM}^2 \\
V_{LF} &= ASC_{LF} + \beta_{Fcost} cost_{LF} \\
V_{EF} &= ASC_{EF} + \beta_{Fcost} cost_{EF} \\
V_{MF} &= ASC_{MF} + \beta_{Fcost} cost_{MF}.
\end{aligned}
$$

Power series estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{BM}	-0.563	0.147	-3.83
2	ASC_{LF}	-0.162	0.370	-0.44
3	ASC_{EF}	-0.377	0.814	-0.46
4	ASC_{MF}	0.215	0.532	0.40
5	β_{Mcost1}	-0.227	0.0427	-5.32
6	β_{Mcost2}	0.000475	0.0000936	5.07
7	β_{Fcost}	-0.107	0.0218	-4.91

Summary statistics
Number of observations $= 434$
$\mathcal{L}(0) = -560.250$
$\mathcal{L}(\hat{\beta}) = -475.465$
$\bar{\rho}^2 = 0.139$

Table 4: Estimation results for the power series expansion

From the estimation results presented in Table 4, it may be noted that the coefficient of the squared term is positive while the coefficient of the linear term is negative, and the coefficient of the linear term is greater in absolute value than that of the squared term. However, since the squared term is very small in magnitude, the total effect is expected to remain negative in the cost range which can be easily verified through a plot of utility versus cost.

To test whether or not we should prefer the power series expansion specification over the linear specification, we need to perform a likelihood ratio test.

The null hypothesis in this case is:

$$H_0 : \beta_{\text{Mcost2}} = 0$$

The χ^2 statistic for this null hypothesis is as follows:

$$-2(L(\hat{\beta}_R) - L(\hat{\beta}_U)) = -2(-476.040 + 475.465) = 1.150$$
$$\chi^2_{0.95,1} = 3.841 > 1.150$$

where now the unrestricted model (U) corresponds to the power series specification. Therefore, we can not reject the null hypothesis of a linear specification at a 95 % level of confidence, and we select the linear specification over the power series expansion specification.

The Box-Cox Transformation

Files to use with Biogeme:	
Model file:	*SpecTest_Tel_boxcox.py*
Data file:	*telephone.dat*

In this section, we analyze the possibility of testing non-linear transformations of variables which are non-linear in the unknown parameters. One such transformation is the Box-Cox expressed as

$$\frac{x^\lambda - 1}{\lambda}, \text{ where } x \geq 0.$$

where λ is a parameter that has to be estimated. We apply such a transformation to the measured cost variable. The utilities remain the same with the substitution of the measured cost variable with its Box-Cox transformation. A snapshot showing such a transformation is in Figure 2.

The parameter λ is estimated along with the other parameters.

The estimation results are shown in Table 5. The estimate of λ was not found to be statistically significantly different from 0. However, it is statistically significantly different from 1 (t-statistic w.r.t. 1 is -2.51). Therefore, we should prefer this non-linear specification over the linear specification. We can also perform a likelihood ratio test as follows. The null hypothesis is given by:

```
#Utilities
V_BM = ASC_BM * one + B_MCOST * (cost1**LAMBDA-1)/LAMBDA
V_SM = ASC_SM * one + B_MCOST * (cost2**LAMBDA-1)/LAMBDA
V_LF = ASC_LF * one + B_FCOST * cost3
V_EF = ASC_EF * one + B_FCOST * cost4
V_MF = ASC_MF * one + B_FCOST * cost5
```

Figure 2: Biogeme snapshot for the Box-Cox transformation

Box-Cox estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	$\mathrm{ASC}_{\mathrm{BM}}$	-0.695	0.166	-4.19
2	$\mathrm{ASC}_{\mathrm{LF}}$	-1.76	1.20	-1.46
3	$\mathrm{ASC}_{\mathrm{EF}}$	-1.98	1.39	-1.43
4	$\mathrm{ASC}_{\mathrm{MF}}$	-1.39	1.28	-1.09
5	β_{Fcost}	-0.104	0.0215	-4.83
6	β_{Mcost}	-1.30	0.880	-1.47
7	λ	0.234	0.305	0.77

Summary statistics
Number of observations = 434
$\mathcal{L}(0) = -560.250$
$\mathcal{L}(\hat{\beta}) = -472.624$
$\bar{\rho}^2 = 0.144$

Table 5: Estimation results for the Box-Cox transformation

$$H_0 : \lambda = 1$$

The χ^2 statistic for this null hypothesis is as follows:

$$-2(L(\hat{\beta}_L) - L(\hat{\beta}_{BC})) = -2(-476.040 + 472.624) = 6.832$$
$$\chi^2_{0.95,1} = 3.841 < 6.832$$

Therefore, the null hypothesis of a linear specification can be rejected at a 95 % level of confidence, and we prefer the Box-Cox transformation.

Swissmetro Case

Market Segmentation

Files to use with Biogeme:	
Model files:	*SpecTest_SM_male.py*,
	SpecTest_SM_female.py,
	SpecTest_SM_full.py,
Data file:	*swissmetro.dat*

In this example, the segmentation is made on the gender variable. We first create two market segments as follows:

- Male: all observations where MALE=1 belong to this subgroup.
- Female: all observations where MALE=0 belong to this subgroup.

Following the procedure described in Ben-Akiva and Lerman (1985) (pages 194-204), we estimate a model on the full data set. Then we run the same model for each gender group separately. Note that we make use of the `[#Exclude]` section in the model specification file to define which observations should be excluded for the estimation. We obtain the values shown in Table 6. The expressions of the utility functions are the same for all models. Note that we define the dummy variable SENIOR which takes the value 1 for individuals with age above 65 and 0 otherwise.

$$
\begin{aligned}
V_{car} &= \mathrm{ASC_{car}} + \beta_{\mathrm{time}}\mathrm{CAR\_TT} + \beta_{\mathrm{car\_cost}}\mathrm{CAR\_CO} + \beta_{\mathrm{senior}}\mathrm{SENIOR} \\
V_{train} &= \beta_{\mathrm{time}}\mathrm{TRAIN\_TT} + \beta_{\mathrm{train\_cost}}\mathrm{TRAIN\_COST} + \beta_{\mathrm{he}}\mathrm{TRAIN\_HE} + \\
&\quad \beta_{\mathrm{ga}}\mathrm{GA} \\
V_{SM} &= \mathrm{ASC_{SM}} + \beta_{\mathrm{time}}\mathrm{SM\_TT} + \beta_{\mathrm{SM\_cost}}\mathrm{SM\_COST} + \beta_{\mathrm{he}}\mathrm{SM\_HE} + \\
&\quad \beta_{\mathrm{senior}}\mathrm{SENIOR} + \beta_{\mathrm{ga}}\mathrm{GA}
\end{aligned}
$$

The null hypothesis is of no taste variation across the market segments:

$$H_0 \; : \; \beta_{Male} = \beta_{Female}$$

Note that in the above equation Male and Female refer to market segments and not to variables in the dataset.

Model	Log likelihood	Number of coefficients
Male	-3680.002	9
Female	-1110.618	9
Restricted model	-4927.167	9

Table 6: Values for the market segmentation test

The likelihood ratio test (with 18-9=9 degrees of freedom) yields

$$\begin{aligned} \mathrm{LR} &= -2(L_N(\hat{\beta}) - \sum_{g=1}^{G} L_{N_g}(\hat{\beta}^g)) \\ &= -2(-4927.167 + 3680.002 + 1110.618) = 273.094 \\ \chi^2_{0.95,9} &= 16.920 \end{aligned}$$

and we can therefore reject the null hypothesis at a 95% level of confidence.

Test of Non-Nested Hypotheses

Files to use with Biogeme:
Model files: *SpecTest_SM_M1.py*, *SpecTest_SM_M2.py*, *SpecTest_SM_MC.py*
Data file: *swissmetro.dat*

In discrete choice analysis, we often perform tests based on the so-called nested hypotheses, which means that we specify two models such that the first one (the restricted model) is a special case of the second one (the unrestricted model). For this type of comparison, the classical likelihood ratio test can be applied. However, there are situations in which we aim at comparing models which are not nested, meaning that one model cannot be obtained as a restricted version of the other. One way to compare two non-nested models is to build a composite model from which both models can be derived. We can thus perform two likelihood ratio tests for each of the restricted models against the composite model. This procedure is known as the Cox test of separate families of hypothesis.

Cox Test

The Cox test is described in detail in Ben-Akiva and Lerman (1985), pages 171-174, and in the Textbook of the course, in section "Tests of Non-Nested Hypothesis". Assume that we want to test a model M_1 against another model M_2 (and one model is not a restricted version of the other). We start by generating a composite model M_C such that both models M_1 and M_2 are restricted cases of M_C. We then test M_1 against M_C and M_2 against M_C using the likelihood ratio test. There are three possible outcomes of this test:

- One of the two models is rejected. Then we keep the one that is not rejected.
- Both models are rejected. Then better models should be developed. The composite model could be used as a new basis for future specifications.
- Both models are accepted. Then we choose the model with the higher $\bar{\rho}^2$ index.

We show next the expressions of the utility functions used for the three different models M_1, M_2 and M_C. M_1 has the following systematic utilities

$$\begin{aligned}
V_{car} &= ASC_{car} + \beta_{car_time} CAR_TT + \beta_{car_cost} CAR_CO \\
V_{train} &= \beta_{train_time} TRAIN_TT + \beta_{train_cost} TRAIN_CO \\
V_{SM} &= ASC_{SM} + \beta_{SM_time} SM_TT + \beta_{SM_cost} SM_CO
\end{aligned}$$

where both the time and cost related coefficients are *alternative specific*. The systematic utilities of M_2 are

$$\begin{aligned}
V_{car} &= ASC_{car} + \beta_{time} CAR_TT + \beta_{car_cost} CAR_CO \\
V_{train} &= \beta_{time} TRAIN_TT + \beta_{train_cost} TRAIN_CO + \\
&\quad \beta_{he} TRAIN_HE + \beta_{ga} GA \\
V_{SM} &= ASC_{SM} + \beta_{time} SM_TT + \beta_{SM_cost} SM_CO + \beta_{he} SM_HE \\
&\quad + \beta_{ga} GA
\end{aligned}$$

where only the cost related coefficient is assumed to be alternative specific, headway of train and SM has been added, and one socio-economic variable has been added to the model. We now define the composite model M_C with the following systematic utilities

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{car\_time} CAR\_TT + \beta_{car\_cost} CAR\_CO \\
V_{train} &= \beta_{train\_time} TRAIN\_TT + \beta_{train\_cost} TRAIN\_CO + \\
&\quad \beta_{he} TRAIN\_HE + \beta_{ga} GA \\
V_{SM} &= ASC_{SM} + \beta_{SM\_time} SM\_TT + \beta_{SM\_cost} SM\_CO + \\
&\quad \beta_{he} SM\_HE + \beta_{ga} GA
\end{aligned}
$$

Models used for the Cox test		
Model	Parameters	Description
M_1	8	two ASC's, three alternative specific *time* coefficients and three alternative specific *cost* coefficients
M_2	8	two ASC's, one generic *time* coefficient, three alternative specific *cost* coefficients, one generic *headway* coefficient and one socio-economic coefficient
M_C	10	two ASC's, three alternative specific *time* coefficients, three alternative specific *cost* coefficients, one generic *headway* coefficient and one socio-economic coefficient

Table 7: Summary of the different model specifications

In Table 7, we summarize the differences between the various models, and we show in Tables 8, 9 and 10 the estimation results for the M_1, M_2 and M_C models, respectively.

At this point, we can apply the likelihood ratio test for M_1 against M_C. In this case, the null hypothesis is:

$$H_0 : \beta_{he} = \beta_{ga} = 0$$

As usual, $-2(L(M_1) - L(M_C))$ is χ^2 distributed with $K = 2$ degrees of freedom. In this case, we have:

M_1 model: estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.260	0.138	-1.89
2	ASC_{SM}	0.113	0.106	1.06
3	β_{car_cost}	-0.00785	0.00149	-5.26
4	β_{train_cost}	-0.0308	0.00193	-15.98
5	β_{SM_cost}	-0.0113	0.000790	-14.24
6	β_{car_time}	-0.0129	0.00163	-7.91
7	β_{train_time}	-0.00870	0.00118	-7.34
8	β_{SM_time}	-0.0112	0.00178	-6.25

Summary statistics
Number of observations $= 6759$
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5065.901$
$\bar{\rho}^2 = 0.271$

Table 8: Estimation results for the M_1 model

M_2 model: estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.872	0.140	-6.24
2	ASC_{SM}	-0.410	0.103	-3.99
3	β_{car_cost}	-0.00934	0.00116	-8.02
4	β_{train_cost}	-0.0284	0.00176	-16.08
5	β_{SM_cost}	-0.0104	0.000743	-13.99
6	β_{time}	-0.0111	0.00120	-9.22
7	β_{he}	-0.00533	0.00102	-5.25
8	β_{ga}	0.521	0.191	2.72
Summary statistics				
Number of observations = 6759				
$\mathcal{L}(0) = -6958.425$				
$\mathcal{L}(\hat{\beta}) = -5055.843$				
$\bar{\rho}^2 = 0.272$				

Table 9: Estimation results for the M_2 model

M_C model: estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.529	0.158	-3.35
2	ASC_{SM}	-0.126	0.116	-1.08
3	β_{car_cost}	-0.00776	0.00150	-5.18
4	β_{train_cost}	-0.0300	0.00200	-14.97
5	β_{SM_cost}	-0.0108	0.000828	-12.99
6	β_{car_time}	-0.0129	0.00162	-7.94
7	β_{train_time}	-0.00866	0.00120	-7.22
8	β_{SM_time}	-0.0111	0.00179	-6.19
9	β_{he}	-0.00535	0.00101	-5.31
10	β_{ga}	0.513	0.193	2.65

Summary statistics
Number of observations $= 6759$
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5047.205$
$\bar{\rho}^2 = 0.273$

Table 10: Estimation results for the M_C model

$$-2(-5065.901 + 5047.205) = 37.392 > 5.991$$

The result of this first test is that we can reject the null hypothesis. Applying the same test for M_2 against M_C, we have

$$H_0 : \beta_{car\_time} = \beta_{train\_time} = \beta_{SM\_time}.$$

In this case, the likelihood ratio test with $K = 2$ degrees of freedom gives

$$-2(-5055.843 + 5047.215) = 17.276 > 5.991$$

and we can therefore reject the null hypothesis in this case as well. Since both models are rejected, better models should be developed. If both models were accepted, we would choose the one with the higher $\bar{\rho}^2$ index.

Tests of Non-Linear Specifications

Files to use with Biogeme:	
Model files:	*SpecTest_SM_piecewise.py*,
	SpecTest_SM_powerseries.py,
	SpecTest_SM_boxcox.py
Data file:	*swissmetro.dat*

In the previous case study, the models were specified with linear in parameter formulations of the deterministic parts of the utilities (i.e. parameters that remain constant throughout the whole range of the values of each variable). However, in some cases non-linear specifications may be more justified. In this section, we test three different non-linear specifications of the deterministic utility functions (see Ben-Akiva and Lerman, 1985, pages 174-179). Namely, piecewise linear approximation, power series method and Box-Cox transformation are used below.

Piecewise Linear Approximation

In this first example, we want to test the hypothesis that the value of the travel time related parameter for the train alternative assumes different values for different ranges of values of the variable itself. We split the

```
    #Expressions
TRAIN_TT1  = DefineVariable('TRAIN_TT1',min( TRAIN_TT , 90))
TRAIN_TT2  = DefineVariable('TRAIN_TT2',max(0,min( TRAIN_TT - 90, 90)))
TRAIN_TT3  = DefineVariable('TRAIN_TT3',max(0,min( TRAIN_TT - 180, 90)))
TRAIN_TT4  = DefineVariable('TRAIN_TT4',max(0, TRAIN_TT - 270))
```

Figure 3: Biogeme snapshot concerning the piecewise variables definition

range of values for travel time t (which is $t \in [35, 1022]$, expressed in minutes) into four different intervals: $train_{tt1} \in [0, 90]$, $train_{tt2} \in]90, 180]$, $train_{tt3} \in]180, 270]$ and $train_{tt4} > 270$. We show in Figure 3 the corresponding Biogeme code.

The systematic utility expressions used in this model are

$$\begin{aligned}
V_{car} &= ASC_{car} + \beta_{car_time} CAR_TT + \beta_{car_cost} CAR_CO \\
V_{train} &= \beta_{train_time1} TRAIN_TT1 + \beta_{train_time2} TRAIN_TT2 + \\
&\quad \beta_{train_time3} TRAIN_TT3 + \beta_{train_time4} TRAIN_TT4 + \\
&\quad \beta_{train_cost} TRAIN_CO + \beta_{he} TRAIN_HE + \beta_{GA} GA \\
V_{SM} &= ASC_{SM} + \beta_{SM_time} SM_TT + \beta_{SM_cost} SM_CO + \beta_{he} SM_HE + \\
&\quad \beta_{GA} GA
\end{aligned}$$

We can see from the estimation results shown in Table 11 that all time coefficients related to the piecewise linear expression are negative. The coefficient associated with very long trips is the largest in magnitude in an absolute sense, meaning that trips longer than 4 hours and a half are more penalizing the utility function of the train alternative.

We perform the likelihood ratio test where the restricted model is the one with linear train travel time (the M_C model from the previous section) and the unrestricted model is the piecewise linear specification. The χ^2 statistic for the null hypothesis is given by

$$H_0 : \beta_{train\_time1} = \beta_{train\_time2} = \beta_{train\_time3} = \beta_{train\_time4}$$

The test yields

$$-2(-5047.205 + 5041.952) = 10.506$$

Piecewise linear model: estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	$\mathsf{ASC}_{\mathsf{car}}$	-0.991	0.434	-2.28
2	$\mathsf{ASC}_{\mathsf{SM}}$	-0.584	0.421	-1.39
3	$\beta_{\mathsf{car\_cost}}$	-0.00776	0.00150	-5.18
4	$\beta_{\mathsf{train\_cost}}$	-0.0301	0.00204	-14.78
5	$\beta_{\mathsf{SM\_cost}}$	-0.0107	0.000828	-12.97
6	$\beta_{\mathsf{car\_time}}$	-0.0129	0.00162	-7.94
7	$\beta_{\mathsf{train\_time1}}$	-0.0135	0.00508	-2.65
8	$\beta_{\mathsf{train\_time2}}$	-0.0109	0.00180	-6.05
9	$\beta_{\mathsf{train\_time3}}$	-0.00208	0.00224	-0.93
10	$\beta_{\mathsf{train\_time4}}$	-0.0179	0.00551	-3.25
11	$\beta_{\mathsf{SM\_time}}$	-0.0112	0.00179	-6.24
12	β_{he}	-0.00534	0.00101	-5.30
13	β_{ga}	0.515	0.193	2.67

Summary statistics
Number of observations = 6759
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5041.952$
$\bar{\rho}^2 = 0.274$

Table 11: Estimation results for the piecewise linear model

and since $\chi^2_{0.95,3} = 7.815$, we can reject the null hypothesis of a linear train travel time at a 95% level of confidence.

The Power Series Expansion

We introduce here a power series expansion for the train travel time variable. In principle, we could add a polynomial expression but here we introduce just the squared term. The subsequent model specification is practically the same as the M_C model, with the exception of the train alternative:

$$\begin{aligned} V_{train} = & \beta_{train\_time} \mathsf{TRAIN\_TT} + \beta_{train\_time\_sq} \mathsf{TRAIN\_TT\_SQ} + \\ & \beta_{train\_cost} \mathsf{TRAIN\_CO} + \beta_{he} \mathsf{TRAIN\_HE} + \\ & \beta_{GA} \mathsf{GA} \end{aligned}$$

The estimation results for this specification are shown in Table 12. The estimated parameter associated with the linear term of the power series expansion is negative while the estimated parameter associated with the squared term is positive. However, the cumulative effect of the travel time variable on the utility is still negative, as can be easily verified by a plot of utility versus travel time for a reasonable range of rail travel time.

We perform the likelihood ratio test where the restricted model is the one with linear train travel time (the M_C model from the previous section) and the unrestricted model is the power series expansion specification. The χ^2 statistic for the null hypothesis is given by:

$$H_0 : \beta_{train\_time^2} = 0$$

The test yields

$$-2(-5047.205 + 5046.573) = 1.264$$

and since $\chi^2_{0.95,1} = 3.841$, we can not reject the null hypothesis of a linear rail travel time at a 95% level of confidence.

Power series model: estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.693	0.190	-3.65
2	ASC_{SM}	-0.289	0.149	-1.94
3	β_{car_cost}	-0.00776	0.00150	-5.18
4	β_{train_cost}	-0.0299	0.00201	-14.86
5	β_{SM_cost}	-0.0108	0.000828	-12.99
6	β_{car_time}	-0.0129	0.00162	-7.95
7	β_{train_time}	-0.0109	0.00190	-5.72
8	$\beta_{train\_time\_sq}$	0.00000628	0.00000282	2.23
9	β_{SM_time}	-0.0111	0.00178	-6.23
10	β_{he}	-0.00537	0.00101	-5.31
11	β_{ga}	0.515	0.194	2.65

Summary statistics
Number of observations $= 6759$
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5046.573$
$\bar{\rho}^2 = 0.273$

Table 12: Estimation results for the power series model

```
V_train = ASC_SBB * one + B_TRAIN_COST * TRAIN_COST + B_HE * TRAIN_HE +
+ B_GA * GA + B_TRAIN_TIME * ( ( ( TRAIN_TT ** LAMBDA  ) - 1 ) /  LAMBDA )
```

Figure 4: Biogeme snapshot of Box-Cox transformation

The Box-Cox Transformation

In this section, we analyze the possibility of testing non-linear transformations of variables that are non-linear in the unknown parameters. One possible transformation is the Box-Cox, expressed as

$$\frac{x^\lambda - 1}{\lambda}, \text{ where } x \geq 0.$$

We apply this transformation to the train time variable. The utilities remain exactly the same, with the substitution of such a variable with its Box-Cox transformation. This introduces one more unknown parameter, λ. We show in Figure 4 a Biogeme snapshot from the model specification file to visualize how the non-linear in parameters utility functions are implemented. We present the utility function of the train alternative.

The results related to the Box-Cox transformed model are shown in Table 13. The Box-Cox transformation reduces to a linear function as a special case when the parameter λ is equal to 1. Looking at the estimated values, we see that λ is significantly different from 1 at a 95 % level of confidence (t-stat = -2.13). Note though that the parameter $\beta_{\text{train\_time}}$ associated with train travel time is not significant.

We can also perform a likelihood ratio test as follows. The null hypothesis is given by:

$$H_0 : \lambda = 1$$

The χ^2 statistic for this null hypothesis is as follows:

$$-2(L(\hat{\beta}_L) - L(\hat{\beta}_{BC})) = -2(-5047.205 + 5045.420) = 3.570$$
$$\chi^2_{0.95,1} = 3.841 > 3.570$$

Therefore, the null hypothesis of a linear specification can not be rejected at a 95 % level of confidence. Note that the t-test and the likelihood ratio test for testing one restriction are asymptotically equivalent. Here the t-stat with respect to 1 is equal to -2.13, so λ is close to being insignificant (w.r.t. 1). In small samples, the likelihood ratio test is preferred to the t-test. Therefore, we prefer the linear specification over the Box-Cox transformation in this case.

Box-Cox transformed model: estimation results

Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-1.72	1.01	-1.71
2	ASC_{SM}	-1.32	1.01	-1.31
3	β_{car_cost}	-0.00776	0.00150	-5.18
4	β_{train_cost}	-0.0298	0.00200	-14.90
5	β_{SM_cost}	-0.0107	0.000828	-12.98
6	β_{car_time}	-0.0129	0.00162	-7.95
7	β_{train_time}	-0.128	0.160	-0.80
8	β_{SM_time}	-0.0111	0.00178	-6.23
9	β_{he}	-0.00535	0.00101	-5.30
10	β_{ga}	0.508	0.194	2.62
11	λ	0.465	0.251	1.85

Summary statistics

Number of observations = 6759

$\mathcal{L}(0) = -6958.425$

$\mathcal{L}(\hat{\beta}) = -5045.420$

$\bar{\rho}^2 = 0.273$

Table 13: Estimation results for the Box-Cox transformed model

Part 2: Multivariate Extreme Value Models

1 Objectives of Part 2 of the Case Study

The topic of Part 2 of this case study is the specification and estimation of Multivariate Extreme Value (MEV) models. Different specifications are introduced using a stepwise modeling strategy, increasing the complexity at each step. The objectives of this case study can be summarized as follows:

- Specification and estimation of Nested Logit (NL) models.
- Testing of the nesting parameters.
- Estimation of Cross Nested Logit (CNL) models, with fixed alpha parameters.
- Estimation of CNL models with unknown alpha parameters.

2 Theoretical Reminder

The family of the Multivariate Generalized Extreme Value models has been introduced by Mc- Fadden (1978). The underlying common factor for all the MEV models is the fact that the random parts of the utility functions for all the alternatives are jointly extreme value distributed. The MNL model represents a specific instance of the MEV family under the assumption that the correlations between the alternatives are all equal to zero. Other much used models belonging to the MEV family are the Nested Logit and the Cross Nested Logit models. We analyze the derivation of both of them in the remainder of this section.

2.1 The Family of MEV Models

In its more general formulation, the expression of the probability of choosing alternative i within the choice set C for a certain decision maker is given by:

$$P(i|C) = \frac{y_i G_i(y_1, ..., y_J)}{\mu G(y_1, ..., y_J)} = \frac{e^{V_i + \ln G_i(...)}}{\sum_{j=1}^{J} e^{V_j + \ln G_j(...)}} \tag{1}$$

where J represents the number of alternatives, $y_i = e^{V_i}$ where V_i is the systematic utility for alternative i and $G_i = \frac{\partial G}{\partial y_i}$. One important fact arises from equation (1):

The utilities of the alternatives are function not only of their own attributes but also of the attributes of competing alternatives, through the partial derivative of the generating G function (Ben-Akiva and Lerman (1985)). The flexibility of the MEV models arises from the possibility of obtaining different correlation structures varying the functional form of the function G which has to satisfy the following four conditions:

- G is non-negative;
- G is homogeneus of degree $\mu > 0$, that is $G(\alpha y) = \alpha^{\mu} G(y)$;
- $\lim_{y_i \to +\infty} G(y_1, ..., y_i, ..., y_J) = +\infty$, for each $i = 1, ..., J$;
- the k^{th} partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even, i.e. for any distinct indices $i_1, ..., i_k \in 1, ..., J$, we have

$$(-1)^k \frac{\partial^k G}{\partial y_{i_1} ... \partial y_{i_k}}(x) \leq 0, \forall y \in \mathbb{R}^J_+$$

Assuming for the generating function G the form

$$G(y) = \sum_{j \in C} y_j^{\mu}$$

we obtain the MNL model.

2.2 The NL Model

In those cases where is reasonable to partition the choice set into subsets, called *nests*, the Nested Logit model can be derived from the general MEV family. The partitioning of the choice set reflects the fact that those alternatives belonging to the same nest share some unobserved attributes and are therefore correlated among themselves (see Ben-Akiva and Lerman (1985), Train (2003), and Wen and Koppelman (2001)). As a consequence, for any two alternatives in the same nest, the ratio of probabilities is independent of the attributes (or existence) of all other alternatives. If two alternatives belong to different nests, the probability ratio can depend on the other alternatives. In other words, the IIA hypothesis holds inside the nest but not between alternatives over different nests. A first intuitive way to derive the NL model is through a sequential approach. Given a partition of the available choice set C into $m = 1, ..., M$ non-overlapping nests and assuming alternative i belongs to nest C_m, we define the probability of choosing alternative i given C as follows:

$$P(i|C) = P(i|C_m)P(C_m|C)$$

The first term represents the probability to choose alternative i given that an alter- native in nest C_m has been chosen while the second term represents the marginal probability of choosing an alternative in nest C_m given the available choice set C. The intuition behind the NL model consists of looking at this formulation as two logit models. The first refers to the choice of nest C_m in the available choice set C. The second refers to the choice of alternative i within the nest C_m. In this case, the systematic utility for nest C_m is composed of two terms: the first depending on variables which are *nest specific*, and the second representing the expected extra utility that decision maker n receives from the choice among those alternatives belonging to the nest C_m. This concept has been formalized by Ben-Akiva (1973) with the introduction of the *inclusive value* (or *log sum*). For more theoretical details about the sequential estimation procedure and the NL model in general, we refer interested readers to Ben-Akiva and Lerman (1985) and Train (2003). We conclude this section with a more general derivation of the NL model. We obtain the NL correlation structure starting from the MEV formulation and using the following generating function, assuming the existence of M different nests:

$$G(y) = \sum_{m=1}^{M} \left(\sum_{j \in C_m} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

where C_m is the set of alternatives inside nest m, with $m = 1, ..., M$. The structural parameter $\mu_m > 1$ and the correlation between the utilities belonging to nest m is given by $1 - \frac{1}{\mu_m^2}$.

2.3 The CNL Model

In the NL formulation, one alternative can be assigned to only one nest. We would say that, given a partition of the choice set into M nests, an alternative j has a degree of membership to nest k (α_{jk}) equal to 1 while $\alpha_{jm} = 0, \forall m \neq k$ with $m = 1, ..., k, ..., M$. Intuitively speaking, it is possible that an alternative shares unobserved attributes from different nests, showing characteristics from all of them or more than one at the same time. As a consequence, it is not possible to associate it with uniquely one nest. This situation is modeled assuming that alternative j has different degrees of membership to different nests, i.e. the cross-correlation between different nests is not zero. The CNL model can be derived from the general MEV formulation assuming the following generating function:

$$G(y_1, ..., y_J) = \sum_{m=1}^{M} \left(\sum_{j \in C_m} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$

where α_{jm} is the degree of membership of alternative j to nest m. For more theoretical details on the CNL model, we refer the reader to Bierlaire (2001), Wen and Koppelman (2001), Vovsha (1997), and Small (1987).

3 Datasets

Same as in Part 1, you can choose between the *Residential Telephone Services*, and the *Swissmetro* datasets. A detailed description of each dataset can be found in Appendix A. Examples of model specifications are provided for both datasets in the following sections.

4 Your tasks

The procedure you should follow to work on the case study can be summarized as follows:

- based on your experience with the case study formulate a hypothesis on possible nesting structure(s);

- perform the estimation of a nested logit model. Give an interpretation of the obtained results and check the validity of the NL model;
- try to modify the model specification, improving the goodness-of-fit and keeping coefficient estimates coherent with the behavioral hypothesis;
- specify a cross nested structure and estimate the cross nested logit model with fixed and variable alpha parameters. Give an interpretation of the obtained results and check the validity of the CNL model;
- perform and discuss relevant statistical tests at each level.

5 Report Content

Your final report should contain:

- the presentation of your best model specification;
- the results of some tests used to arrive at your best model specification;
- a discussion of the estimated parameter values;
- all the final BIOGEME outputs for all discussed models.

Choice of Residential Telephone Services Case

Estimation of a Nested Logit Model

Files to use with Biogeme:
Model file: MEV_Tel_NL_unrestricted.py, MEV_Tel_NL_restricted.py
Data file: telephone.dat

We start by giving some examples of possible nesting structures for the Nested Logit (NL) model in Figure 5. See Chapter 10 in Ben-Akiva and Lerman (1985) for details on the NL model.

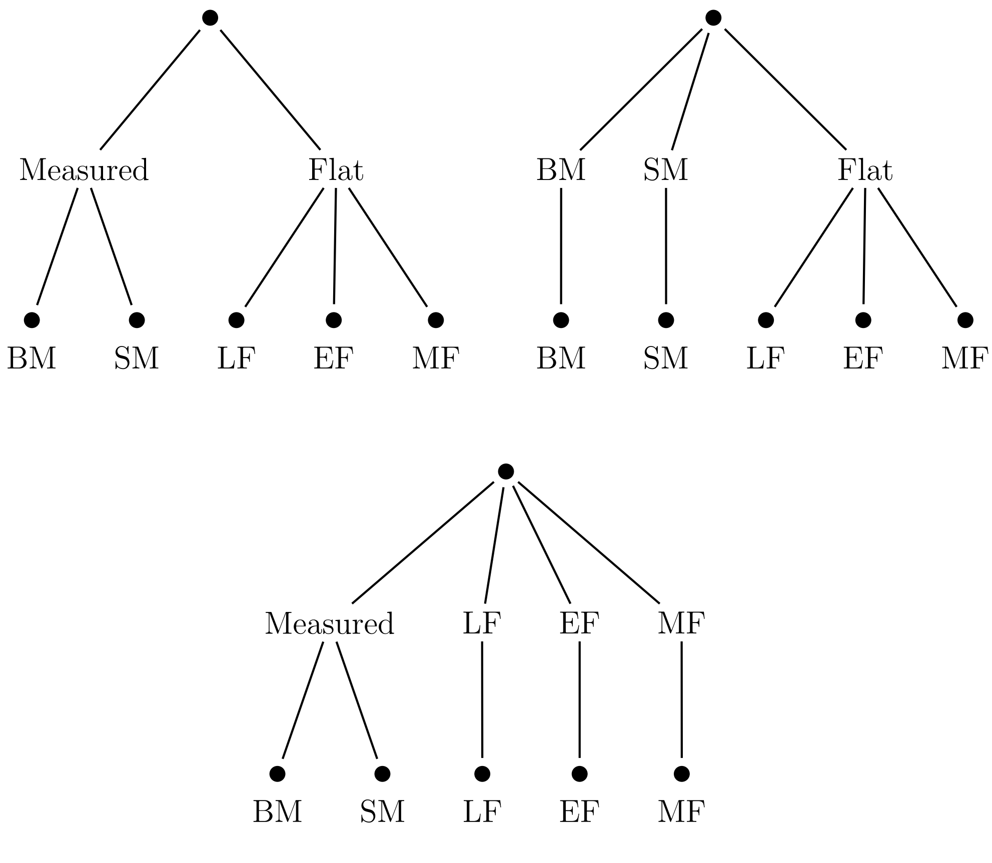

Figure 5: Possible nesting structures

```
# parameters relevant to the nests (name for report, starting value,
#  lower bound, upper bound, 0:estimate / 1:keep it fixed)
MU_FLAT = Beta('MU_FLAT',1,1,10, 0)
MU_MEAS = Beta('MU_MEAS',1,1,10, 0)

#Definitions of nests
N_FLAT = MU_FLAT, [3, 4, 5]
N_MEAS = MU_MEAS, [1, 2]

nests = N_FLAT, N_MEAS
```

Figure 6: Biogeme snapshot

The sample model file describes the first nesting structure shown in Figure 5. The expressions of the utilities for this simple NL model are

$$\begin{aligned}
V_{BM} &= \mathrm{ASC}_{BM} + \beta_{cost} \ln(\mathrm{cost}_{BM}) \\
V_{SM} &= \beta_{cost} \ln(\mathrm{cost}_{SM}) \\
V_{LF} &= \mathrm{ASC}_{LF} + \beta_{cost} \ln(\mathrm{cost}_{LF}) \\
V_{EF} &= \mathrm{ASC}_{EF} + \beta_{cost} \ln(\mathrm{cost}_{EF}) \\
V_{MF} &= \mathrm{ASC}_{MF} + \beta_{cost} \ln(\mathrm{cost}_{MF}).
\end{aligned}$$

We show a snapshot of the Biogeme code in Figure 6. The nests are named and defined as nests, and we write the alternatives that belong to each. Here the alternative numbers must correspond to those used in the utilities section of the command file. The estimation results of the NL model are shown in Table 14.

To be consistent with random utility theory, the inequality $\frac{\mu}{\mu_m} < 1$ with μ being normalized to 1 implies $\mu_m > 1$. To see if this is the case here, we can test the null hypothesis $H_0 : \mu_{meas} = \mu_{flat} = 1$. Since there are multiple restrictions here, we cannot do multiple t-tests. We should do a likelihood ratio test as follows. The test statistic for the null hypothesis is given by

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-477.557 + 473.219) = 8.676$$

NL with generic attributes					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t stat. 0*	Robust *t stat. 1*
1	$\mathrm{ASC}_{\mathsf{BM}}$	-0.378	0.117	-3.22	
2	$\mathrm{ASC}_{\mathsf{LF}}$	0.893	0.158	5.64	
3	$\mathrm{ASC}_{\mathsf{EF}}$	0.847	0.391	2.17	
4	$\mathrm{ASC}_{\mathsf{MF}}$	1.41	0.238	5.90	
5	β_{cost}	-1.49	0.243	-6.12	
6	μ_{meas}	2.06	0.573	3.60	1.86
7	μ_{flat}	2.29	0.764	3.00	1.69

Summary statistics
Number of observations = 434
$\mathcal{L}(0) = -560.250$
$\mathcal{L}(\hat{\beta}) = -473.219$
$\bar{\rho}^2 = 0.143$

Table 14: NL with generic attributes

```
# parameters relevant to the nests (name for report, starting value,
#  lower bound, upper bound, 0:estimate / 1:keep it fixed)
MU_FLAT = Beta('MU_FLAT',1,1,10,0)
MU_MEAS = MU_FLAT
```

Figure 7: Biogeme snapshot

where the restricted model is the logit model (*MNL_Tel_generic.py*) and the unrestricted model is the nested logit model. The test statistic is asymptotically χ^2 distributed with 2 degrees of freedom since there are 2 restrictions. Since $8.676 > 5.991$ (the critical value of the χ^2 distribution with 2 degrees of freedom at a 95 % level of confidence), we reject the null hypothesis (logit model) and accept the nested logit model.

The μ_m's of the two nests can be set equal to each other too. This can be done by constraining the two nest coefficients to be equal (Figure 7).

The estimation results for this last specification are shown in Table 15.

NL with linear constraints					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t stat.* 0	Robust *t stat.* 1
1	ASC_{BM}	-0.368	0.113	-3.26	
2	ASC_{LF}	0.882	0.154	5.74	
3	ASC_{EF}	0.833	0.401	2.08	
4	ASC_{MF}	1.39	0.232	5.96	
5	β_{cost}	-1.50	0.243	-6.18	
6	μ_{meas}	2.16	0.563	3.84	2.06
7	μ_{flat}	2.16	0.563	3.84	2.06

Summary statistics
Number of observations = 434
$\mathcal{L}(0) = -560.250$
$\mathcal{L}(\hat{\beta}) = -473.288$
$\bar{\rho}^2 = 0.145$

Table 15: NL with linear constraints on nest parameters

Estimation of a Cross-Nested Logit Model with Fixed Alphas

Files to use with Biogeme:
Model file: MEV_Tel_CNL_fix.py
Data file: telephone.dat

In this section and the next one, we specify two different Cross-Nested Logit (CNL) models (see Abbe et al. (2007) for a detailed description of the CNL model) using both fixed and variable degrees of membership. The major premise here is that such specifications are mainly for demonstration purposes. However, an assumption that might make sense is that the local flat alternative (LF) is likely to be correlated with both measured and flat options. Like the measured plans, the LF plan is a reasonable option for users with relatively basic needs. So, the LF option may belong to both nests. Based on this hypothesis, the proposed cross-nested structure is shown in Figure 8.

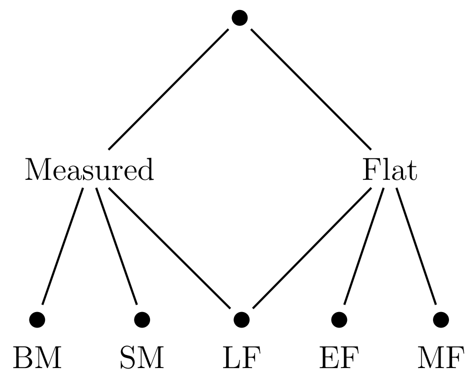

Figure 8: The cross-nested structure

We present the CNL model with the same deterministic utility functions as in the previous model. The nest parameters are constrained to 2.16, their values from the restricted NL estimation, to allow clearer focus on the cross-nested estimation structure. The corresponding snapshot from the Biogeme code for this cross-nesting specification is shown in Figure 9.

Note that we define α_{CNL} so that the LF alternative belongs equally to both the flat and the measured nests. This assumption will be relaxed in the next section. Thus, CNL with fixed α's is a restricted model of CNL with variable α's. The estimation results are shown in Table 16.

Cross-Nested Logit Model with Variable Alphas

Files to use with Biogeme:
Model file: MEV_Tel_CNL_var.py
Data file: telephone.dat

In the previous CNL model, we assumed that the LF alternative belongs equally to the measured nest and the flat nest by fixing $\alpha_{\mathsf{LF\_meas}}$ and $\alpha_{\mathsf{LF\_flat}}$ to be equal to 0.5. This assumption can be relaxed, and we can estimate the share of LF in each nest during the estimation of the model parameters. The corresponding Biogeme snapshot is shown in Figure 10. From the results presented in Table 17, we see that the alternative LF has a very small share in the measured nest.

```
# parameters relevant to the nests
MU_FLAT = Beta('MU_FLAT',2.16,1,10,1)
MU_MEAS = Beta('MU_MEAS',2.16,1,10,1)
#
a_FLAT_LF = Beta('a_FLAT_LF',0.5,0,1,1)
a_MEAS_LF = Beta('a_MEAS_LF',0.5,0,1,1)
a_FLAT_EF = Beta('a_FLAT_EF',1,0,1,1)
a_FLAT_MF = Beta('a_FLAT_MF',1,0,1,1)
a_MEAS_BM = Beta('a_MEAS_BM',1,0,1,1)
a_MEAS_SM = Beta('a_MEAS_SM',1,0,1,1)
#Definitions of nests
alpha_N_FLAT = {1: 0, 2: 0, 3: a_FLAT_LF, 4: a_FLAT_EF, 5: a_FLAT_MF}
alpha_N_MEAS = {1: a_MEAS_BM, 2: a_MEAS_SM, 3: a_MEAS_LF, 4: 0, 5: 0}

nest_N_FLAT = MU_FLAT, alpha_N_FLAT
nest_N_MEAS = MU_MEAS, alpha_N_MEAS

nests = nest_N_FLAT, nest_N_MEAS
```

Figure 9: Biogeme snapshot

CNL with α_{CNL} fixed				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t stat. 0*
1	ASC_{BM}	-0.356	0.0730	-4.87
2	ASC_{LF}	0.867	0.0870	9.96
3	ASC_{EF}	0.465	0.395	1.18
4	ASC_{MF}	0.791	0.161	4.90
5	β_{cost}	-1.24	0.132	-9.41
Summary statistics				
Number of observations = 434				
$\mathcal{L}(0) = -612.574$				
$\mathcal{L}(\hat{\beta}) = -480.146$				
$\bar{\rho}^2 = 0.208$				

Table 16: CNL with α_{CNL} fixed

We also want to underline the fact that in both CNL specifications the condition

$$\sum_{m} \alpha_{jm} = 1$$

has been imposed. Such a condition is not necessary for the validity of the model. It is imposed for identification purposes. We refer the interested reader to Abbe et al. (2007) for more theoretical details.

To select between the restricted nested logit and CNL model with variable α's, we can test the null hypothesis $H_0 : \alpha_{LF\_flat} = 1$. Since there is a single restriction, we can use either a t-test or a likelihood ratio test which are asymptotically equivalent. The t-statistic with respect to 1 is -0.0807, which indicates that α_{LF_flat} is not significantly different from 1, and hence we accept the null hypothesis (restricted nested logit model) and reject the CNL model with variable α's.

We can also do a likelihood ratio test as follows. The test statistic for the null hypothesis is given by

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-473.288 + 473.250) = 0.076$$

where the restricted model is the nested logit model and the unrestricted model is the CNL model. The test statistic is asymptotically χ^2 distributed with 1 degree of freedom since there is 1 restriction. Since $0.076 < 3.841$ (the critical value of the χ^2 distribution with 1 degree of freedom at a 95 % level of confidence), we accept the null hypothesis (restricted nested logit model) and reject the CNL model with variable α's. We can thus conclude that the LF alternative is correlated only with the flat nest but not with the measured nest.

To select between the CNL model with fixed α's and the CNL model with variable α's, we can test the null hypothesis $H_0 : \alpha_{LF\_flat} = 0.5$. Since there is a single restriction, we can use either a t-test or a likelihood ratio test which are asymptotically equivalent. The t-statistic with respect to 0.5 is 5.04, which indicates that α_{LF_flat} is significantly different from 0.5, and hence we reject the null hypothesis (CNL model with fixed α's) and accept the CNL model with variable α's.

We can also do a likelihood ratio test as follows. The test statistic for the null hypothesis is given by

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-480.146 + 473.250) = 13.792$$

```
# parameters relevant to the nests
MU_FLAT = Beta('MU_FLAT',2.16,1,10,1)
MU_MEAS = Beta('MU_MEAS',2.16,1,10,1)
#
a_FLAT_LF = Beta('a_FLAT_LF',0.5,0,1,0)
a_FLAT_EF = Beta('a_FLAT_EF',1,0,1,1)
a_FLAT_MF = Beta('a_FLAT_MF',1,0,1,1)
a_MEAS_SM = Beta('a_FLAT_SM',1,0,1,1)
a_MEAS_BM = Beta('a_MEAS_BM',1,0,1,1)
#
a_MEAS_LF = 1 - a_FLAT_LF
```

Figure 10: Biogeme snapshot

where the restricted model is the CNL model with fixed α's and the unrestricted model is the CNL model with variable α's. The test statistic is asymptotically χ^2 distributed with 1 degree of freedom since there is 1 restriction. Since $13.792 > 3.841$ (the critical value of the χ^2 distribution with 1 degree of freedom at a 95 % level of confidence), we reject the null hypothesis (CNL model with fixed α's) and accept the CNL model with variable α's.

Since the restricted nested logit model is preferred to the CNL model with variable α's, and the CNL model with variable α's is preferred to the CNL model with fixed α's, we select the restricted nested logit model over the CNL models.

CNL with α_{CNL} variable					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t stat. 0*	Robust *t stat. 1*
1	$\mathrm{ASC}_{\mathrm{BM}}$	-0.368	0.0759	-4.85	
2	$\mathrm{ASC}_{\mathrm{LF}}$	0.943	0.128	7.39	
3	$\mathrm{ASC}_{\mathrm{EF}}$	0.827	0.395	2.09	
4	$\mathrm{ASC}_{\mathrm{MF}}$	1.37	0.192	7.17	
5	β_{cost}	-1.49	0.155	-9.65	
6	$\alpha_{\mathrm{LF\_flat}}$	0.931	0.0855	10.89	-0.807
Summary statistics					
Number of observations $= 434$					
$\mathcal{L}(0) = -612.574$					
$\mathcal{L}(\hat{\beta}) = -473.250$					
$\bar{\rho}^2 = 0.218$					

Table 17: CNL with α_{CNL} variable

Swissmetro Case

Estimation of a Nested Logit Model

Files to use with Biogeme:
Model file: *MEV_SM_NL.py*
Data file: *swissmetro.dat*

We start with a Nested Logit (NL) specification, where the car and train alternatives are both assigned to the same nest and the Swissmetro is alone in a second nest, as shown in Figure 12. See Chapter 10 in Ben-Akiva and Lerman (1985) for details on the NL model.

The expressions of the systematic utility functions for each alternative used in this model specification are

$$\begin{aligned}
V_{\mathsf{car}} &= \mathrm{ASC}_{\mathsf{car}} + \beta_{\mathsf{CAR_time}}\mathsf{CAR_TT} + \beta_{\mathsf{cost}}\mathsf{CAR_CO} \\
V_{\mathsf{train}} &= \beta_{\mathsf{TRAIN_time}}\mathsf{TRAIN_TT} + \beta_{\mathsf{cost}}\mathsf{TRAIN_CO} + \beta_{\mathsf{he}}\mathsf{TRAIN_HE} + \\
&\quad \beta_{\mathsf{GA}}\mathsf{GA} \\
V_{\mathsf{sm}} &= \mathrm{ASC}_{\mathsf{SM}} + \beta_{\mathsf{SM_time}}\mathsf{SM_TT} + \beta_{\mathsf{cost}}\mathsf{SM_CO} + \beta_{\mathsf{he}}\mathsf{SM_HE} \\
&\quad \beta_{\mathsf{GA}}\mathsf{GA},
\end{aligned}$$

and in Figure 11 an extract from the *.py* file illustrating the nest specification with Biogeme is shown. Note that only one of the two nest parameters can be estimated. The estimation results are shown in Table 18.

The alternative specific constants show a preference for the Swissmetro alternative compared to the other modes, all the rest remaining constant. The cost and travel time coefficients have the expected negative sign. The coefficient related to the ownership of the Swiss annual season ticket (GA) is positive as expected, reflecting the preference for the SM and train alternatives with respect to the car alternative. The negative estimated value of the headway parameter β_{he} indicates that the higher the headway, the lower the frequency of service, and thus the lower the utility. Finally, the scale parameter of the random term associated with the *classic* nest has been estimated as $\mu_{\mathsf{classic}} = 2.33$.

To be consistent with random utility theory, the inequality $\frac{\mu}{\mu_m} < 1$ with μ

```
# parameters relevant to the nests
MU_classic = Beta('MU_classic',1,1,10,0)

# Definition of nests
innovative = 1.0, [2]
classic = MU_classic, [1, 3]

nests = classic, innovative
```

Figure 11: Biogeme snapshot

NL model					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t-stat. 0*	Robust *t-stat. 1*
1	$\mathrm{ASC}_{\mathsf{car}}$	-0.274	0.0875	-3.13	
2	$\mathrm{ASC}_{\mathsf{SM}}$	0.0259	0.104	0.25	
3	β_{cost}	-0.00718	0.000598	-12.01	
4	$\beta_{\mathsf{car\_time}}$	-0.00741	0.00117	-6.33	
5	$\beta_{\mathsf{train\_time}}$	-0.0104	0.00107	-9.68	
6	$\beta_{\mathsf{SM\_time}}$	-0.00807	0.00168	-4.79	
7	β_{he}	-0.00352	0.000661	-5.32	
8	β_{ga}	0.719	0.103	6.97	
9	μ_{classic}	2.33	0.176	13.29	7.56

Summary statistics
Number of observations $= 6759$
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5146.325$
$\bar{\rho}^2 = 0.259$

Table 18: NL estimation results

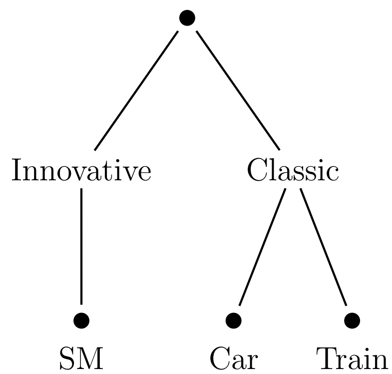

Figure 12: The correlation structure of the specified NL model

being normalized to 1 implies $\mu_m > 1$. To see if this is the case here, we can test the null hypothesis $H_0 : \mu_m = 1$. Since there is a single restriction, we can use either a t-test or a likelihood ratio test which are asymptotically equivalent. The t-statistic with respect to 1 can be computed as follows: $\frac{(\hat{\mu}_m - 1)}{\text{std err of } \hat{\mu}_m}$. Here the t-statistic with respect to 1 is 7.56, which indicates that $\mu_{classic}$ is significantly different from 1, and hence there is a significant correlation between the car and train alternatives.

We can also do a likelihood ratio test as follows. The test statistic for the null hypothesis is given by

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-5272.203 + 5146.325) = 251.756$$

where the restricted model is the corresponding logit model (*MNL_SM.py*) and the unrestricted model is the nested logit model. The test statistic is asymptotically χ^2 distributed with 1 degree of freedom since there is 1 restriction. Since $251.756 > 3.841$ (the critical value of the χ^2 distribution with 1 degree of freedom at a 95 % level of confidence), we reject the null hypothesis (logit model) and accept the nested logit model.

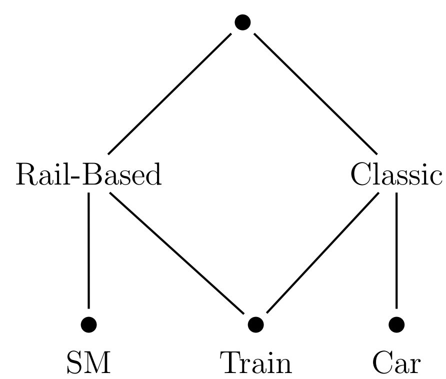

Figure 13: A representative scheme for the CNL correlation structure.

Estimation of a Cross-Nested Logit Model with Fixed Alphas

Files to use with Biogeme:
model file: MEV_SM_CNL_fix.py
data file: swissmetro.dat

In this model, we relax the assumption that an alternative can belong to only one nest and we assume that the train alternative can be assigned to two different nests. This correlation structure is motivated by considering the train alternative as a *classic* transportation mode (along with the car against the more innovative Swissmetro) on one hand, and as a rail-based mode (as the Swissmetro) on the other hand. We represent this cross-nested structure in Figure 13. See Abbe et al. (2007) for a detailed description of the Cross-Nested Logit (CNL) model.

In Figure 14 we show a snapshot from the Biogeme `.py` file illustrating the CNL nest specification. The estimation results are shown in Table 19. The alternative-specific constants now have a negative sign. All other coefficients have the expected signs.

In this CNL specification, we have fixed the $\alpha_{\mathrm{train\_classic}}$ and $\alpha_{\mathrm{train\_rail}}$ coefficients to 0.5. It means that we assume that the train alternative equally belongs to both nests *classic* and *rail-based*. This assumption will be relaxed in the next section. Thus, CNL with fixed α's is a restricted model of CNL

CNL model with fixed α's					
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t-stat. 0*	Robust *t-stat. 1*
1	$\text{ASC}_{\mathsf{car}}$	-0.838	0.0787	-10.65	
2	ASC_{SM}	-0.457	0.0744	-6.15	
3	β_{cost}	-0.00705	0.000526	-13.39	
4	$\beta_{\mathsf{car\_time}}$	-0.00628	0.00122	-5.17	
5	$\beta_{\mathsf{train\_time}}$	-0.00863	0.00105	-8.18	
6	$\beta_{\mathsf{SM\_time}}$	-0.00715	0.00151	-4.74	
7	β_{he}	-0.00298	0.000533	-5.58	
8	β_{ga}	0.618	0.0940	6.57	
9	μ_{classic}	2.85	0.260	10.93	7.09
10	$\mu_{\mathsf{rail\_based}}$	4.73	0.483	9.78	7.71
Summary statistics					
Number of observations = 6759					
$\mathcal{L}(0) = -6958.425$					
$\mathcal{L}(\hat{\beta}) = -5120.738$					
$\bar{\rho}^2 = 0.263$					

Table 19: Estimation results for the CNL specification. The α coefficients are fixed.

```
 # parameters relevant to the nests
Rail_based = Beta('Rail_based',1,1,10,0)
classic = Beta('classic',1,1,10,0)
Rail_based_SM = Beta('Rail_based_SM',1,1e-05,1,1)
Rail_based_Train = Beta('Rail_based_Train',0.5,1e-05,1,1)
classic_Car = Beta('classic_Car',1,1e-05,1,1)
classic_Train = 1 - Rail_based_Train

# Definition of nests
alpha_Rail_based = {1: Rail_based_Train, 2: Rail_based_SM, 3: 0}
alpha_classic = {1: classic_Train, 2: 0, 3: classic_Car}

nest_Rail_based = Rail_based, alpha_Rail_based
nest_classic = classic, alpha_classic

nests = nest_Rail_based, nest_classic
```

Figure 14: Biogeme snapshot

with variable α's.

Estimation of a Cross-Nested Logit Model with Unknown Alphas

Files to use with Biogeme:
Model file: MEV_SM_CNL_var.py
Data file: swissmetro.dat

In Table 20, we show the results for the CNL specification with variable α coefficients. We also want to underline the fact that in both CNL specifications the condition

$$\sum_m \alpha_{jm} = 1$$

has been imposed. Such a condition is not necessary for the validity of the model. It is imposed for identification purposes. We refer the interested reader to Abbe et al. (2007) for more theoretical details.

To select between the nested logit and CNL model with variable α's, we can

CNL model with unknown α's					
Parameter number	Parameter name	Parameter estimate	standard error	*t-stat. 0*	*t-stat. 1*
1	ASC_{car}	-0.849	0.0692	-12.26	
2	ASC_{SM}	-0.460	0.0656	-7.01	
3	β_{cost}	-0.00697	0.000440	-15.85	
4	β_{car_time}	-0.00621	0.000583	-10.66	
5	β_{train_time}	-0.00849	0.000660	-12.85	
6	β_{SM_time}	-0.00711	0.000745	-9.54	
7	β_{he}	-0.00293	0.000510	-5.75	
8	β_{ga}	0.620	0.0886	7.00	
9	$\mu_{classic}$	2.87	0.212	13.54	8.82
10	μ_{rail_based}	4.90	0.722	6.78	5.40
11	$\alpha_{train\_classic}$	0.486	0.0265	18.35	-19.40
12	α_{train_rail}	0.514	0.0265	19.40	-18.35

Summary statistics
Number of observations = 6759
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5120.608$
$\bar{\rho}^2 = 0.262$

Table 20: Estimation results for the CNL specification. The α coefficients are estimated.

test the null hypothesis $H_0 : \alpha_{train\_rail} = 0, \mu_{rail\_based} = 1$. Since there are multiple restrictions, we cannot use multiple t-tests but should rather use a likelihood ratio test as follows. The test statistic for the null hypothesis is given by

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-5146.325 + 5120.608) = 51.434$$

where the restricted model is the nested logit model and the unrestricted model is the CNL model with variable α's. The test statistic is asymptotically χ^2 distributed with 2 degrees of freedom since there are 2 restrictions. Since $51.434 > 5.991$ (the critical value of the χ^2 distribution with 2 degrees of freedom at a 95 % level of confidence), we reject the null hypothesis (nested logit model) and accept the CNL model with variable α's. We can thus conclude that the train alternative is correlated with both Swissmetro and car alternatives.

To select between the CNL model with fixed α's and the CNL model with variable α's, we can test the null hypothesis $H_0 : \alpha_{train\_rail} = 0.5$. Since there is a single restriction, we can use either a t-test or a likelihood ratio test which are asymptotically equivalent. The t-statistic with respect to 0.5 is 0.53, which indicates that α_{train_rail} is not significantly different from 0.5, and hence we accept the null hypothesis (CNL model with fixed α's) and reject the CNL model with variable α's.

We can also do a likelihood ratio test as follows. The test statistic for the null hypothesis is given by

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-5120.738 + 5120.608) = 0.260$$

where the restricted model is the CNL model with fixed α's and the unrestricted model is the CNL model with variable α's. The test statistic is asymptotically χ^2 distributed with 1 degree of freedom since there is 1 restriction. Since $0.260 < 3.841$ (the critical value of the χ^2 distribution with 1 degree of freedom at a 95 % level of confidence), we accept the null hypothesis (CNL model with fixed α's) and reject the CNL model with variable α's.

As a conclusion, since both the nested logit model and the CNL model with fixed α's are restricted models of the CNL model with variable α's, and since we have rejected the nested logit model and accepted the CNL model with fixed α's, we select the CNL model with fixed α's.

References

Abbe, E., Bierlaire, M. and Toledo, T. (2007). Normalization and correlation of cross-nested logit models, *Transportation Research B: Methodological* **41**(7): 795–808.

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.

Bierlaire, M., Axhausen, K. and Abay, G. (2001). The acceptance of modal innovation: The case of swissmetro, *Proceedings of the 1st Swiss Transportation Research Conference*, Ascona, Switzerland. www.strc.ch/bierlaire.pdf.

Train, K., McFadden, D. and Ben-Akiva, M. (1987). The demand for local telephone service: a fully discrete model of residential calling patterns and service choices, *Rand Journal of Economics* .

A Datasets

Residential Telephone Services Data Set

Context

Local telephone service typically involves the choice between flat (i.e., a fixed monthly charge for unlimited calls within a specified geographical area) and measured (i.e., a reduced fixed monthly charge for a limited number of calls and additional usage charges for additional calls) services. Various flat rate services differ by the size of the geographical area within which calling is provided at no extra charge, the monthly charge being higher for larger areas. Measured services differ with respect to the threshold number (or dollar value) of calls beyond which the customer is charged. The availability of each service may depend on the geographical location within the service area.

In developing a model of the residential demand for local telephone service, it is necessary to explicitly account for the inter-relationship between class of service choice and usage patterns. For example, expected usage patterns will influence the household's choice of service option since households with high usage levels typically could minimize their monthly bill for local telephone service by choosing some sort of flat rate service, while households with relatively low usage would be better off with a measured service. Given that a household has chosen a particular service option, usage patterns would be dependent to a certain extent upon the service option that is chosen since it determines the marginal price of calls. To accommodate these interrelationships, the model representing the household's choice of calling patterns and service options needs to include:

1. choice of the service option, which is modeled conditional upon the calling portfolio chosen by the household;

2. choice of the calling portfolio or the usage pattern as represented by the number and duration of calls by time of day and calling band.

This case study deals only with the first choice.

Data Collection

A household survey was conducted in 1984 for a telephone company among 434 households in Pennsylvania. The dataset involves choices among five calling plans and consists of various attributes and socio-economic characteristics. It was originally used to develop a model system to predict residential telephone demand (Train et al., 1987).

Variables and Descriptive Statistics

In the current application, five types of services are involved: two measured options and three flat options. The availability of these service options varies depending upon geographic location. Table 21 below lists the five service alternatives and their availability within the different service areas. Names and definitions of the variables are shown in Table 22. Some descriptive statistics of the dataset are summarized in Table 23.

Complications caused by very few respondents choosing alternative 4: If you examine the dataset, you see that only 3 of the respondents chose alternative 4 (extended area flat service). This implies that it is not possible to estimate numerous alternative specific coefficients for alternative 4. The intuition is that the dataset does not provide enough information on why people chose or did not choose alternative 4. If you try to estimate too many alternative specific coefficients for alternative 4, you get "Singularity in the Hessian" error, and in order to estimate the model you have to reduce the number of coefficients specific to alternative 4. A practical solution to this problem is to use an "enriched sample" although such a sample is not available here. It is however not recommended to omit the observations for which the chosen alternative is 4 or combine alternative 4 with a different alternative.

		Availability		
Service option	Description	metro, suburban, some perimeter areas	other perimeter areas	non-metro areas
1. Budget measured	No fixed monthly charge; usage charges apply to each call made.	yes	yes	yes
2. Standard measured	A fixed monthly charge covers up to a specified dollar amount (greater than the fixed charge) of local calling, after which usage charges apply to each call made.	yes	yes	yes
3. Local flat	A greater monthly charge that may depend upon residential location; unlimited free calling within local calling area; usage charges apply to calls made outside local calling area.	yes	yes	yes
4. Extended area flat	A further increase in the fixed monthly charge to permit unlimited free calling within an extended area.	no	yes	no
5. Metro area flat	The greatest fixed monthly charge that permits unlimited free calling within the entire metropolitan area.	yes	yes	no

Table 21: Service options and their availability

Name	**Description**
age0 age1 age2 age3 age4 age5 age6 age7	number of household members under age 6 number of household members age 6-12 number of household members age 13-19 number of household members age 20-29 number of household members age 30-39 number of household members age 40-54 number of household members age 55-64 number of household members 65 and older
area	location of household residence 1=metro, 2=suburban, 3=perimeter with extended, 4=perimeter without extended, 5=non-metro
avail1, avail2, avail3, avail4, avail5	binary indicators of availability of each option. availX=0 if alternative X is not available to the household, availX=1 if alternative X is available to the household
choice	chosen alternative (dependent variable) 1=budget measured, 2=standard measured, 3=local flat, 4=extended flat, 5=metro flat
cost1, cost2, cost3, cost4, cost5	costX = monthly cost (in $) of alternative X.
employ	number of household members employed
inc	annual household income 1=under $10,000, 2=$10,000-20,000, 3=$20,000-30,000, 4=$30,000-40,000, 5=0ver $40,000
ones	ones = 1 for all observations
status	marital status 1=single, 2=married, 3=widowed, 4=divorced, 5=other
users	number of phone users in household

Table 22: Description of variables

	mean	max	min	stand dev	range
age0	0.21	4	0	0.53	4
age1	0.23	3	0	0.58	3
age2	0.24	4	0	0.67	4
age3	0.41	3	0	0.71	3
age4	0.44	2	0	0.73	2
age5	0.36	2	0	0.67	2
age6	0.31	3	0	0.61	3
age7	0.38	2	0	0.65	2
area	2.93	5	1	1.65	4
avail1	1.00	1	1	0.00	0
avail2	1.00	1	1	0.00	0
avail3	1.00	1	1	0.00	0
avail4	0.03	1	0	0.17	1
avail5	0.65	1	0	0.48	1
choice	2.65	5	1	1.17	4
cost1	11.73	433.5	3.28	24.13	430.22
cost2	11.49	432.8	5.78	23.90	427.02
cost3	14.82	435.5	7.03	23.56	428.47
cost4	62.19	433.03	10.48	117.88	422.55
cost5	27.48	38.28	23.28	4.17	15
employ	1.07	3	0	0.89	3
inc	2.53	5	1	1.28	4
ones	1.00	1	1	0.00	0
status	2.22	5	1	0.91	4
users	2.30	6	1	1.28	5

Table 23: Descriptive Statistics

Swissmetro Data Set

This dataset consists of survey data collected on the trains between St. Gallen and Geneva, Switzerland, during March 1998. The respondents provided information in order to analyze the impact of the modal innovation in transportation, represented by the Swissmetro, a revolutionary mag-lev underground system, against the usual transport modes represented by car and train.

Context

Innovation in the market for intercity passenger transportation is a difficult enterprise as the existing modes: private car, coach, rail as well as regional and long-distance air services continue to innovate in their own right by offering new combinations of speeds, services, prices and technologies. Consider for example high-speed rail links between the major centers or direct regional jet services between smaller countries. The Swissmetro SA in Geneva is promoting such an innovation: a mag-lev underground system operating at speeds up to 500 km/h in partial vacuum connecting the major Swiss conurbations, in particular along the Mittelland corridor (St. Gallen, Zurich, Bern, Lausanne and Geneva).

Data Collection

The Swissmetro is a true innovation. It is therefore not appropriate to base forecasts of its impact on observations of existing revealed preferences (RP) data. It is necessary to obtain data from surveys of hypothetical markets/situations, which include the innovation, to assess the impact. Survey data were collected on rail-based travels, interviewing 470 respondents. Due to data problems, only 441 are used here. Nine stated choice situations were generated for each of 441 respondents, offering three alternatives: rail, Swissmetro and car (only for car owners).

A similar method for relevant car trips with a household or telephone survey was deemed impractical. The sample was therefore constructed using license plate observations on the motorways in the corridor by means of video recorders. A total of 10529 relevant license plates were recorded during September 1997. The central Swiss car license agency had agreed to send up to 10000 owners of these cars a survey-pack. Until April 1998, 9658 letters were mailed, of which 1758 were returned. A total of 1070 persons filled in the survey completely and were willing to participate in the second SP survey, which was generated using the same approach used for the rail interviews. 750 usable SP surveys were returned, from the license-plate based survey.

Variables and Descriptive Statistics

The variables of the dataset are described in Tables 24 and 25, and the descriptive statistics are summarized in Table 26. A more detailed description of the data set as well as the data collection procedure is given in Bierlaire et al. (2001).

Variable	Description
GROUP	Different groups in the population. 2: current rail users, 3: current road users
SURVEY	Equivalent to GROUP but using different coding: 0: train users, 1: car users
SP	It is fixed to 1 (stated preference survey)
ID	Respondent identifier
PURPOSE	Travel purpose. 1: Commuter, 2: Shopping, 3: Business, 4: Leisure, 5: Return from work, 6: Return from shopping, 7: Return from business, 8: Return from leisure, 9: other
FIRST	First class traveler (0 = no, 1 = yes)
TICKET	Travel ticket. 0: None, 1: Two way with half price card, 2: One way with half price card, 3: Two way normal price, 4: One way normal price, 5: Half day, 6: Annual season ticket, 7: Annual season ticket Junior or Senior, 8: Free travel after 7pm card, 9: Group ticket, 10: Other
WHO	Who pays (0: unknown, 1: self, 2: employer, 3: half-half)
LUGGAGE	0: none, 1: one piece, 3: several pieces
AGE	It captures the age class of individuals. The age-class coding scheme is of the type: 1: $\text{age} \leq 24$, 2: $24 < \text{age} \leq 39$, 3: $39 < \text{age} \leq 54$, 4: $54 < \text{age} \leq 65$, 5: $65 < \text{age}$, 6: not known
MALE	Traveler's Gender 0: female, 1: male
INCOME	Traveler's income per year [thousand CHF] 0 or 1: under 50, 2: between 50 and 100, 3: over 100, 4: unknown
GA	Variable capturing the effect of the Swiss annual season ticket for the rail system and most local public transport. It is 1 if the individual owns a GA, zero otherwise.
ORIGIN	Travel origin (a number corresponding to a Canton, see Table 27)

Table 24: Description of variables

Variable	Description
DEST	Travel destination (a number corresponding to a Canton, see Table 27)
TRAIN_AV	Train availability dummy
CAR_AV	Car availability dummy
SM_AV	SM availability dummy
TRAIN_TT	Train travel time [minutes]. Travel times are door-to-door making assumptions about car-based distances (1.25*crow-flight distance)
TRAIN_CO	Train cost [CHF]. If the traveler has a GA, this cost equals the cost of the annual ticket.
TRAIN_HE	Train headway [minutes] Example: If there are two trains per hour, the value of TRAIN_HE is 30.
SM_TT	SM travel time [minutes] considering the future Swissmetro speed of 500 km/h
SM_CO	SM cost [CHF] calculated at the current relevant rail fare, without considering GA, multiplied by a fixed factor (1.2) to reflect the higher speed.
SM_HE	SM headway [minutes] Example: If there are two Swissmetros per hour, the value of SM_HE is 30.
SM_SEATS	Seats configuration in the Swissmetro (dummy). Airline seats (1) or not (0).
CAR_TT	Car travel time [minutes]
CAR_CO	Car cost [CHF] considering a fixed average cost per kilometer (1.20 CHF/km)
CHOICE	Choice indicator. 0: unknown, 1: Train, 2: SM, 3: Car

Table 25: Description of variables

Variable	Min	Max	Mean	St. Dev.
GROUP	2	3	2.63	0.48
SURVEY	0	1	0.63	0.48
SP	1	1	1.00	0.00
ID	1	1192	596.50	344.12
PURPOSE	1	9	2.91	1.15
FIRST	0	1	0.47	0.50
TICKET	1	10	2.89	2.19
WHO	0	3	1.49	0.71
LUGGAGE	0	3	0.68	0.60
AGE	1	6	2.90	1.03
MALE	0	1	0.75	0.43
INCOME	0	4	2.33	0.94
GA	0	1	0.14	0.35
ORIGIN	1	25	13.32	10.14
DEST	1	26	10.80	9.75
TRAIN_AV	1	1	1.00	0.00
CAR_AV	0	1	0.84	0.36
SM_AV	1	1	1.00	0.00
TRAIN_TT	31	1049	166.63	77.35
TRAIN_CO	4	5040	514.34	1088.93
TRAIN_HE	30	120	70.10	37.43
SM_TT	8	796	87.47	53.55
SM_CO	6	6720	670.34	1441.59
SM_HE	10	30	20.02	8.16
SM_SEATS	0	1	0.12	0.32
CAR_TT	0	1560	123.80	88.71
CAR_CO	0	520	78.74	55.26
CHOICE	1	3	2.15	0.63

Table 26: Descriptive statistics

Number	**Canton**
1	ZH
2	BE
3	LU
4	UR
5	SZ
6	OW
7	NW
8	GL
9	ZG
10	FR
11	SO
12	BS
13	BL
14	Schaffhausen
15	AR
16	AI
17	SG
18	GR
19	AG
20	TH
21	TI
22	VD
23	VS
24	NE
25	GE
26	JU

Table 27: Coding of Cantons