EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire - Dr. de Lapparent - Dr. Sharif

Decision-aid Methodologies in Transportation
Spring 2016

Swissmetro

1 Market Segmentation

Files to use with Biogeme:
Model files: *SpecTest_SM_male.mod*,
SpecTest_SM_female.mod,
SpecTest_SM_full.mod,
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 1. 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.

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

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

Table 1: Values for the market segmentation test

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.

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

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

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

2 Test of Non-Nested Hypotheses

Files to use with Biogeme:
Model files: *SpecTest_SM_M1.mod, SpecTest_SM_M2.mod, SpecTest_SM_MC.mod*
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.

2.1 Cox Test

The Cox test is described in detail in the textbook and in the slides 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 2: Summary of the different model specifications

In Table 2, we summarize the differences between the various models, and we show in Tables 3, 4 and 5 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:

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

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 3: 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 4: 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 5: Estimation results for the M_C model

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.

3 Tests of Non-Linear Specifications

Files to use with Biogeme:
Model files: *SpecTest_SM_piecewise.mod*,
SpecTest_SM_powerseries.mod,
SpecTest_SM_boxcox.mod
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.

3.1 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 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 1 the corresponding Biogeme code.

```
[Expressions]
TRAIN_TT1  =  min( TRAIN_TT , 90)
TRAIN_TT2  =  max(0,min( TRAIN_TT - 90, 90))
TRAIN_TT3  =  max(0,min( TRAIN_TT - 180 , 90))
TRAIN_TT4  =  max(0,TRAIN_TT - 270)
```

Figure 1: Biogeme snapshot concerning the piecewise variables definition

The systematic utility expressions used in this model are

$$\begin{aligned} V_{car} &= ASC_{car} + \beta_{car\_time} CAR\_TT + \beta_{car\_cost} CAR\_CO \\ V_{train} &= \beta_{train\_time1} TRAIN\_TT1 + \beta_{train\_time2} TRAIN\_TT2 + \\ &\quad \beta_{train\_time3} TRAIN\_TT3 + \beta_{train\_time4} TRAIN\_TT4 + \\ &\quad \beta_{train\_cost} TRAIN\_CO + \beta_{he} TRAIN\_HE + \beta_{GA} GA \\ V_{SM} &= \mathrm{ASC}_{SM} + \beta_{SM\_time} SM\_TT + \beta_{SM\_cost} SM\_CO + \beta_{he} SM\_HE + \beta_{GA} GA \end{aligned}$$

We can see from the estimation results shown in Table 6 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

$$H_0 : \beta_{train\_time1} = \beta_{train\_time2} = \beta_{train\_time3} = \beta_{train\_time4} \tag{1}$$

The test yields

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

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.

3.2 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

Piecewise linear model: estimation results				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.991	0.434	-2.28
2	ASC_{SM}	-0.584	0.421	-1.39
3	β_{car_cost}	-0.00776	0.00150	-5.18
4	β_{train_cost}	-0.0301	0.00204	-14.78
5	β_{SM_cost}	-0.0107	0.000828	-12.97
6	β_{car_time}	-0.0129	0.00162	-7.94
7	β_{train_time1}	-0.0135	0.00508	-2.65
8	β_{train_time2}	-0.0109	0.00180	-6.05
9	β_{train_time3}	-0.00208	0.00224	-0.93
10	β_{train_time4}	-0.0179	0.00551	-3.25
11	β_{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 6: Estimation results for the piecewise linear model

squared term. The subsequent model specification is practically the same as the M_C model, with the exception of the train alternative:

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

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 7: Estimation results for the power series model

The estimation results for this specification are shown in Table 7. 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

```
[GeneralizedUtilities]
1      B_TRAIN_TIME * ( ( ( TRAIN_TT ) ^ LAMBDA - 1 ) / LAMBDA )
```

Figure 2: Biogeme snapshot of Box-Cox transformation

null hypothesis is given by:

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

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.

3.3 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

$$\frac{x^\lambda - 1}{\lambda}, \text{ where } x \geq 0. \tag{3}$$

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 2 a Biogeme snapshot from the model specification file to visualize how non-linear in parameters utility functions are implemented.

The results related to the Box-Cox transformed model are shown in Table 8. 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 β_{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 8: Estimation results for the Box-Cox transformed model