Appendix 1: The Case of Swissmetro

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 centuries. 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 was 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 sending 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 1, 2, and 3. The descriptive statistics are summarized in Table 4. A more detailed description of the dataset as well as the data collection procedure is given in Bierlaire et al. (2001).

Variable	Description
GROUP	Different groups in the population
SURVEY	Survey performed in train (0) or car (1)
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 traveller (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: age$\leq$24, 2: 24$<$age$\leq$39, 3: 39$<$age$\leq$54, 4: 54$<$age$\leq$ 65, 5: 65 $<$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 3)

Table 1: Description of variables

Variable	Description
DEST	Travel destination (a number corresponding to a Canton, see Table 3)
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] without considering GA.
TRAIN_FR	Train frequency (headway) [minutes] Example: If there are two trains per hour, the value of TRAIN_FR 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 with a fixed factor (1.2) to reflect the higher speed.
SM_FR	SM frequency (headway) [minutes] Example: If there are two Swissmetros per hour, the value of SM_FR 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 2: Description of variables

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 3: Coding of Cantons

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_FR	30	120	70.10	37.43
SM_TT	8	796	87.47	53.55
SM_CO	6	6720	670.34	1441.59
SM_FR	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	0	3	2.15	0.63

Table 4: Descriptive statistics

Model	Log likelihood	Number of coefficients
Male	-3680.00	9
Female	-1110.62	9
Restricted model	-4927.17	9

Table 5: Values for the market segmentation test

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 in the estimation. We obtain the values reported in Table 5. 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.

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

The null hypothesis is of no taste variations 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

$$\begin{aligned} LR &= -2(L_N(\hat{\beta}) - \sum_{g=1}^{G} L_{N_g}(\hat{\beta}^g)) \\ &= -2(-4927.17 + 3680.00 + 1110.62) = 273.10 \\ \chi^2_{0.95,9} &= 16.92 \end{aligned}$$

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

McFadden IIA Test

Files to use with BIOGEME:	
Model file:	*SpecTest_SM_IIA.mod*,
	SpecTest_SM_socioec_bis.mod
Data file:	*swissmetro.dat, swissmetro-IIA.dat*
Excel worksheet:	*swissmetro-IIA.xls*

We are studying the impact of the modal innovation, represented by the Swissmetro, against traditional transport modes represented by car and train. In this spirit, it would seem logical to expect some kind of *relationship* between the traditional alternatives. They are probably correlated, where the source of this correlation might be the presence of unobserved shared attributes between the car and train alternatives. In order to test this assumption, we follow the procedure explained in Ben-Akiva and Lerman (1985) (pages 183-194). That is, first we estimate a MNL model (*SpecTest_SM_socioec_bis.mod*) on the full data set *swissmetro.dat*. We use the estimated values of the parameters to compute in Excel (or using Biosim) the choice probabilities for each observation (individual) and for each alternative. As discussed above, we assume in this case that the subset of alternatives suspected to be correlated is given by $\widehat{C} = \{\text{car,train}\}$. We then compute in Excel the two corresponding auxiliary variables for each observation of the data file to get the file *swissmetro-IIA.xls*, which we export in the Text format file *swissmetro-IIA.dat*. Now we specify a new model (*SpecTest_SM_IIA.mod*) which includes the auxiliary variables in the utility functions associated with train and car. Finally, we estimate this model on this new data file. We show in Table 6 the estimation results.

The focus in this test is not related to the sign of the estimated IIA parameter. What is important is the value of the t-statistic for such a coefficient. β_{IIA} is significantly different from 0 at a 95% level of confidence. This indicates that the IIA property does not hold for the car and train alternatives. This kind of correlation can be captured with GEV models that are treated in the Part 2 of this case study.

MNL for car/train IIA test				
Variable a number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	0.196	0.161	1.22
2	ASC_{sm}	0.489	0.129	3.79
3	β_{cost}	-0.00115	0.000110	-10.51
4	β_{car_time}	-0.0102	0.000974	-10.52
5	β_{train_time}	-0.0114	0.00114	-9.99
6	β_{sm_time}	-0.0112	0.00168	-6.65
7	β_{fr}	-0.00489	0.00110	-4.45
8	β_{ga}	6.29	0.661	9.51
9	β_{IIA}	0.395	0.116	3.40
Summary statistics				
Number of observations $= 6759$				
$\mathcal{L}(0) = -6958.42$				
$\mathcal{L}(\hat{\beta}) = -5229.90$				
$\bar{\rho}^2 = 0.247$				

Table 6: MNL for IIA test

Test of Non-Nested Hypothesis

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 hypothesis, 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 another one. 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.

Composite Model Test

The composite model test is described in detail in Ben-Akiva and Lerman (1985) (pages 171-174). 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 highest $\bar{\rho}^2$ index.

We report in the following 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

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

where only the cost related attribute is assumed to be alternative specific, frequency 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

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

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

Models used for the composite test		
Model	Parameters	Description
M_1	8	two ASC, three alternative specific *time* coefficients and three alternative specific *cost* coefficients
M_2	8	two ASC, one generic *time* coefficient, three alternative specific *cost* coefficients, one generic *frequency* parameter and one socio-economic variable
M_C	10	two ASC, three alternative specific *time* coefficients, three alternative specific *cost* coefficients, one generic *frequency* parameter and one socio-economic variable

Table 7: Summary of the different model specifications

M_1 model: estimation results				
Variable number	Variable name	Coefficient 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 $= 6759$				
$\mathcal{L}(0) = -6958.42$				
$\mathcal{L}(\hat{\beta}) = -5065.90$				
$\bar{\rho}^2 = 0.271$				

Table 8: Estimation results for the M_1 model

M_2 **model: estimation results**				
Variable number	Variable name	Coefficient 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_{fr}	-0.00533	0.00102	-5.25
8	β_{ga}	0.521	0.191	2.72

Summary statistics
Number of observations $= 6759$
$\mathcal{L}(0) = -6958.42$
$\mathcal{L}(\hat{\beta}) = -5055.84$
$\bar{\rho}^2 = 0.272$

Table 9: Estimation results for the M_2 model

M_C **model: estimation results**				
Variable number	Variable name	Coefficient 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_{fr}	-0.00535	0.00101	-5.31
10	β_{ga}	0.513	0.194	2.65

Summary statistics
Number of observations $= 6759$
$\mathcal{L}(0) = -6958.42$
$\mathcal{L}(\hat{\beta}) = -5047.21$
$\bar{\rho}^2 = 0.273$

Table 10: Estimation results for the M_C model

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

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

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

$$-2(-5065.90 + 5047.21) = 37.38 > 5.991$$

The result of this first test is that we can reject the null hypothesis. Applying the same test on 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

$$-2(-5055.84 + 5047.21) = 17.26 > 5.991$$

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.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 part 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. 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 range of values for rail travel time t (which is $t \in [35, 1022]$, expressed in minutes) into four different intervals:

```
[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

$train_{tt1} \in [0, 90]$, $train_{tt2} \in]90, 180]$, $train_{tt3} \in]180, 270]$ and $train_{tt4} > 270$. We show in Figure 1 the corresponding BIOGEME code.

The systematic utility expressions used in this model are

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

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

$$H0 : \beta_{train\_time1} = \beta_{train\_time2} = \beta_{train\_time3} = \beta_{train\_time4}$$

The test yields

$$-2(-5047.21 + 5041.95) = 10.52$$

and since $\chi^2_{3,0.05} = 7.815$, 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

Piecewise linear model: estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.990	0.434	-2.28
2	ASC_{SM}	-0.583	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_{fr}	-0.00534	0.00101	-5.30
13	β_{ga}	0.515	0.193	2.67

Summary statistics

Number of observations = 6759

$\mathcal{L}(0) = -6958.42$

$\mathcal{L}(\hat{\beta}) = -5041.95$

$\bar{\rho}^2 = 0.274$

Table 11: 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:

$$\begin{aligned} V_{train} = \; & \beta_{train\_time} TRAIN\_TT + \beta_{train\_time\_sq} TRAIN\_TT\_SQ + \\ & \beta_{train\_cost} TRAIN\_CO + \beta_{fr} TRAIN\_FR + \\ & \beta_{GA} GA \end{aligned}$$

Power series model: estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-0.684	0.190	-3.59
2	ASC_{SM}	-0.280	0.150	-1.87
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}	-0.0108	0.00191	-5.66
8	$\beta_{train\_time\_sq}$	0.00000612	0.00000287	2.13
9	β_{SM_time}	-0.0111	0.00178	-6.23
10	β_{fr}	-0.0053	0.00101	-5.31
11	β_{ga}	0.510	0.194	2.63
Summary statistics				
Number of observations $= 6759$				
$\mathcal{L}(0) = -6958.42$				
$\mathcal{L}(\hat{\beta}) = -5046.57$				
$\bar{\rho}^2 = 0.273$				

Table 12: Estimation results for the power series model

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 on the utility variable 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

$$H0 : \beta_{train\_time^2} = 0$$

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

Figure 2: BIOGEME snapshot of Box-Cox transformation

The test yields

$$-2(-5047.21 + 5046.57) = 1.28$$

and since $\chi^2_{1,0.05} = 3.841$, we can accept the null hypothesis of a linear rail travel time at a 95% level of confidence.

The Box-Cox Transformation

In this section, we analyze the possibility of testing for 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 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 13. We know that the Box-Cox transformation reduces to a linear function as a special case, when the parameter λ is fixed 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:

$$-2(L(\hat{\beta}_L) - L(\hat{\beta}_{BC})) = -2(-5047.21 + 5045.42) = 3.58$$
$$\chi^2_{0.95,1} = 3.841 > 3.58$$

Therefore, the null hypothesis of a linear specification is accepted 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). Since the parameter β_{train_time} associated with train travel time is not significant, we are unable to capture a significant effect of the train time variable using the Box-Cox transformation. Therefore, we prefer the linear specification over the Box-Cox transformation in this case.

Box-Cox transformed model: estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	-1.73	1.02	-1.70
2	ASC_{SM}	-1.33	1.02	-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}	-0.129	0.162	-0.80
8	β_{SM_time}	-0.0111	0.00178	-6.23
9	β_{fr}	-0.00535	0.00101	-5.30
10	β_{ga}	0.508	0.194	2.62
11	λ	0.463	0.252	1.84
Summary statistics				
Number of observations = 6759				
$\mathcal{L}(0) = -6958.42$				
$\mathcal{L}(\hat{\beta}) = -5045.42$				
$\bar{\rho}^2 = 0.273$				

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