TRANSP-OR

Exercise session 11

The topic of this lab is mixtures of models. Since the estimation of this type of models some-times can take several hours (or days), we provide the results for some models based on the Swissmetro data. For each type of model, analyse the specification. Which are the underlying assumptions? Is the model correctly specified? What interpretations can you draw from the estimation results? It is also interesting to compare the results with the other models estimated during the previous lab sessions.

Consult the BIOGEME user manual for details on how different random distributions are specified.

Heteroskedastic Model

In this first model specification we assume that the ASCs are randomly distributed with mean $\bar{\alpha}_{car}$ and $\bar{\alpha}_{SM}$ and standard deviation σ_{car} and σ_{SM}. We show below the utility expressions and the related BIOGEME code. We normalize with respect to the train alternative and the estimation results are reported in Table 1.

$$\begin{aligned}
V_{car} &= ASC_{car} + \beta_{time}CAR\_TT + \beta_{cost}CAR\_CO \\
V_{train} &= \beta_{time}TRAIN\_TT + \beta_{cost}TRAIN\_CO + \beta_{fr}TRAIN\_FR \\
V_{SM} &= ASC_{SM} + \beta_{time}SM\_TT + \beta_{cost}SM\_CO + \beta_{fr}SM\_FR
\end{aligned}$$

```
[Utilities]
// Id  Name  Avail  linear-in-parameter expression
31 Car_SP  CAR_AV_SP    ASC_CAR_SP [ ASC_CAR_SP_std ] * one +
                        B_TIME * CAR_TT + B_COST * CAR_CO +
                        B_INCOME * INCOME
11 SBB_SP  TRAIN_AV_SP  ASC_SBB_SP * one + B_TIME * TRAIN_TT +
                        B_COST * TRAIN_COST + B_FR * TRAIN_FR
21 SM_SP   SM_AV        ASC_SM_SP [ ASC_SM_SP_std ] * one + B_TIME * SM_TT
                        + B_COST * SM_COST + B_FR * SM_FR +
                        B_INCOME * INCOME
```

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	$\bar{\alpha}_{car}$	0.244	0.107	2.288
2	$\bar{\alpha}_{SM}$	0.844	0.178	4.749
3	σ_{car}	0.099	0.097	1.018
4	σ_{SM}	2.918	0.417	7.004
5	β_{cost}	-0.017	0.002	- 10.939
6	β_{fr}	-0.008	0.001	-5.717
7	β_{time}	-0.016	0.002	-8.657

Summary statistics
Number of draws=100
Number of observations=6768
L(0)=-6964.66
L($\hat{\beta}$)=-5257.98
$\bar{\rho}^2 = 0.245049$

Table 1: Heteroskedastic specification

Error Component Model

We present two different specifications of error component models. Below, the systematic utility expressions and the related BIOGEME code for the first model are shown. The train and SM modes share the random term ζ_{rail}, which is assumed to be normally distributed $\zeta_{rail} \sim N(m_{rail}, \sigma_{rail}^2)$. We estimate the standard deviation σ_{rail} of this error component, while the mean m_{rail} is fixed to zero. The estimation results are reported in Table 2.

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{time}CAR\_TT + \beta_{cost}CAR\_CO \\
V_{train} &= \beta_{time}TRAIN\_TT + \beta_{cost}TRAIN\_CO + \\
&\quad \beta_{fr}TRAIN\_FR + \zeta_{rail} \\
V_{SM} &= ASC_{SM} + \beta_{time}SM\_TT + \beta_{cost}SM\_CO + \beta_{fr}SM\_FR \\
&\quad + \zeta_{rail}
\end{aligned}
$$

```
[Utilities]
// Id  Name  Avail  linear-in-parameter expression
31  Car_SP  CAR_AV_SP    ASC_CAR_SP * one + B_TIME * CAR_TT +
                         B_COST * CAR_CO
11  SBB_SP  TRAIN_AV_SP  ASC_SBB_SP * one + B_TIME * TRAIN_TT +
                         B_COST * TRAIN_COST + B_FR * TRAIN_FR +
                         RAIL [ RAIL_std ] * one
```

```
21  SM_SP   SM_AV           ASC_SM_SP * one + B_TIME * SM_TT + B_COST * SM_COST
                            + B_FR * SM_FR + RAIL [ RAIL_std ] * one
```

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	0.184	0.080	2.301
2	ASC_{SM}	0.449	0.093	4.802
3	β_{cost}	-0.011	0.0006	-15.915
4	β_{fr}	-0.005	0.0009	-5.449
5	β_{time}	-0.0128	0.001	-12.187
6	σ_{rail}	0.153	0.058	2.665

Summary statistics

Number of draws=100

Number of observations=6768

L(0)=-6964.66

L($\hat{\beta}$)=-5314.7

$\bar{\rho}^2 = 0.236905$

Table 2: First Error component specification

In the second model we use a more complex error structure, the specification is presented below and the estimation results are reported in Table 3.

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{time}CAR\_TT + \beta_{cost}CAR\_CO + \zeta_{classic} \\
V_{train} &= \beta_{time}TRAIN\_TT + \beta_{cost}TRAIN\_CO + \beta_{fr}TRAIN\_FR \\
&\quad + \zeta_{rail} + \zeta_{classic} \\
V_{SM} &= ASC_{SM} + \beta_{time}SM\_TT + \beta_{cost}SM\_CO + \\
&\quad \beta_{fr}SM\_FR + \zeta_{rail}
\end{aligned}
$$

```
[Utilities]
// Id  Name  Avail  linear-in-parameter expression
31  Car_SP  CAR_AV_SP    ASC_CAR_SP * one + B_TIME * CAR_TT +
                        B_COST * CAR_CO + CLASSIC [ CLASSIC_std ] * one
11  SBB_SP  TRAIN_AV_SP  ASC_SBB_SP * one + B_TIME * TRAIN_TT +
                        B_COST * TRAIN_COST + B_FR * TRAIN_FR +
                RAIL [ RAIL_std ] * one +
                        CLASSIC [ CLASSIC_std ] * one
21  SM_SP   SM_AV        ASC_SM_SP * one + B_TIME * SM_TT +
                        B_COST * SM_COST + B_FR * SM_FR +
                        RAIL [ RAIL_std ] * one
```

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC_{car}	0.254	0.110	2.319
2	ASC_{SM}	0.865	0.238	3.630
3	β_{cost}	-0.0166	0.0016	-10.048
4	β_{fr}	-0.0076	0.0013	-5.662
5	β_{time}	-0.016	0.0019	-8.116
6	$\sigma_{classic}$	2.862	0.526	5.437
7	σ_{rail}	0.098	0.101	0.974

Summary statistics
Number of draws=100
Number of observations=6768
L(0)=-6964.66
L($\hat{\beta}$)=-5261.82
$\bar{\rho}^2 = 0.244498$

Table 3: Second Error component specification

Random Coefficients

In this specification the unknown parameters are assumed to be randomly distributed over the population. The utility expressions are shown below as well as the related BIOGEME code. The estimation results are reported in Table 4.

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{time}CAR\_TT + \beta_{car\_cost}CAR\_CO \\
V_{train} &= \beta_{time}TRAIN\_TT + \beta_{train\_cost}TRAIN\_CO + \beta_{fr}TRAIN\_FR \\
V_{SM} &= ASC_{SM} + \beta_{time}SM\_TT + \beta_{SM\_cost}SM\_CO + \beta_{fr}SM\_FR
\end{aligned}
$$

```
[Utilities]
// Id  Name  Avail  linear-in-parameter expression
31 Car_SP  CAR_AV_SP   ASC_CAR_SP * one + B_TIME * CAR_TT
                       + B_CAR_COST [ B_CAR_COST_std ] * CAR_CO
11 SBB_SP  TRAIN_AV_SP ASC_SBB_SP * one + B_TIME * TRAIN_TT
                       + B_TRAIN_COST [ B_TRAIN_COST_std ] * TRAIN_COST
                       + B_FR [ B_FR_std ] * TRAIN_FR
21 SM_SP   SM_AV    ASC_SM_SP * one + B_TIME * SM_TT
                       + B_SM_COST [ B_SM_COST_std ] * SM_COST
                       + B_FR [ B_FR_std ] * SM_FR
```

Different distributions
We report here two examples of BIOGEME code to specify a random coefficient model where

		Estimation results		
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust t $statistic$
1	ASC_{car}	-1.470	0.177	-8.295
2	ASC_{SM}	-0.915	0.130	-7.066
3	m_{car_cost}	-0.017	0.004	-4.113
4	σ_{car_cost}	0.009	0.003	2.681
5	m_{train_cost}	-0.059	0.005	-12.135
6	σ_{train_cost}	0.023	0.002	10.939
7	m_{SM_cost}	-0.016	0.002	-7.477
8	σ_{SM_cost}	0.008	0.002	3.992
9	m_{fr}	-0.006	0.001	-5.116
10	σ_{fr}	0.001	0.004	0.245
11	β_{time}	-0.013	0.002	-7.721

Summary statistics
Number of draws=100
Number of observations=6768
L(0)=-6964.66
L($\hat{\beta}$)=-4979.7
$\bar{\rho}^2 = 0.285004$

Table 4: Random coefficient specification

the parameters are log-normally and Johnsons Sb distributed. Recall that, a variable X is log normally distributed if $y = \ln(X)$ is normally distributed. We can easily define in BIOGEME such a distribution by assuming a generic time coefficient to be log-normally distributed.

```
[GeneralizedUtilities]
11 exp( B_TIME [ B_TIME_std ] ) * CAR_TT
21 exp( B_TIME [ B_TIME_std ] ) * TRAIN_TT
31 exp( B_TIME [ B_TIME_std ] ) * SM_TT
```

In the case of Johnsons SB distribution, the functional form is derived using a Logit-like transformation of a Normal distribution, as defined in the following equation

$$\xi = a + (b - a)\frac{e^{\zeta}}{e^{\zeta} + 1} \tag{1}$$

where $\zeta \sim N(\mu, \sigma^2)$. This distribution is very flexible; it is bounded between a and b and its shape can change from a very flat one to a bimodal, changing the parameters of the normal variable. It requires the estimation of four parameters (a, b, μ and σ) and a nonlinear specification, assuming as before, a generic time coefficient following such a distribution.

```
[GeneralizedUtilities]
11 ( A + ( ( B - A ) * ( exp( B_TIME [ B_TIME_std ] )
   / ( exp( B_TIME [ B_TIME_std ] ) + 1 ) ) ) ) * CAR_TT
21 ( A + ( ( B - A ) * ( exp( B_TIME [ B_TIME_std ] )
   / ( exp( B_TIME [ B_TIME_std ] ) + 1 ) ) ) ) * TRAIN_TT
31 ( A + ( ( B - A ) * ( exp( B_TIME [ B_TIME_std ] )
   / ( exp( B_TIME [ B_TIME_std ] ) + 1 ) ) ) ) * SM_TT
```

Mixed GEV Models

In this example we capture the substitution patterns using a Nested Logit model and we allow for some parameters to be randomly distributed over the population.

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{car\_time}CAR\_TT + \beta_{cost}CAR\_CO \\
V_{train} &= \beta_{train\_time}TRAIN\_TT + \beta_{cost}TRAIN\_CO + \beta_{fr}TRAIN\_FR \\
&\quad + \beta_{ga}GA + \beta_{age}AGE \\
V_{SM} &= ASC_{SM} + \beta_{SM\_time}SM\_TT + \beta_{cost}SM\_CO + \beta_{fr}SM\_FR \\
&\quad + \beta_{ga}GA + \beta_{seats}SEATS
\end{aligned}
$$

We specify a nest composed of the alternatives *car* and *train* representing standard transportation modes, while the Swissmetro alternative represents the technological innovation. We further assume a generic cost parameter and three randomly distributed alternative-specific time parameters. Normal distributions are used for the random coefficients, that is,

$$\beta_{car\_time} \sim N(m_{car\_time}, \sigma^2_{car\_time})$$

		Estimation results			
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t stat. 0*	Robust *t stat. 1*
1	ASC_{car}	0.246	0.134	1.835	
2	ASC_{SM}	0.599	0.134	4.459	
3	β_{age}	0.203	0.032	6.401	
4	m_{car_time}	-0.012	0.001	-12.107	
5	β_{cost}	-0.009	0.0008	-11.179	
6	σ_{car_time}	0.004	0.0005	8.560	
7	β_{ga}	0.888	0.135	6.563	
8	β_{fr}	-0.004	0.0008	-5.575	
9	β_{seats}	-0.288	0.096	-2.986	
10	m_{SM_time}	-0.014	0.001	-10.671	
11	σ_{SM_time}	0.009	0.001	6.324	
12	m_{train_time}	-0.015	0.001	-13.623	
13	σ_{train_time}	0.0003	0.0005	0.634	
14	$\mu_{classic}$	2.229	0.176	12.63	6.966

Summary statistics
Number of draws=100
Number of observations=6768
L(0)=-6964.66
L($\hat{\beta}$)=-5052.93
$\bar{\rho}^2 = 0.27449$

Table 5: Mixed Nested Logit estimation results

$$\begin{aligned} \beta_{train\_time} &\sim N(m_{train\_time}, \sigma^2_{train\_time}) \\ \beta_{SM\_time} &\sim N(m_{SM\_time}, \sigma^2_{SM\_time}). \end{aligned}$$

The estimation results are reported in Table 5.

mbi/tr-jn