Appendix A: The Case of Swissmetro

Estimation of a Nested Logit Model

Files to use with BIOGEME:
Model file: *GEV_SM_NL.mod*
Data file: *swissmetro.dat*

The application of the IIA McFadden test in the case study on specification testing revealed that the IIA assumption does not hold between the car and train alternatives. This is an indication of probable correlation between car and train. We start with a Nested Logit (NL) specification, where the car and train alternatives are assigned both to the same nest and the Swissmetro is alone in a second nest, as shown in Figure 8. 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

$$
\begin{aligned}
V_{car} &= ASC_{car} + \beta_{CAR\_time} CAR\_TT + \beta_{cost} CAR\_CO \\
V_{train} &= \beta_{TRAIN\_time} TRAIN\_TT + \beta_{cost} TRAIN\_CO + \beta_{fr} TRAIN\_FR + \\
&\quad \beta_{GA} GA \\
V_{sm} &= ASC_{SM} + \beta_{SM\_time} SM\_TT + \beta_{cost} SM\_CO + \beta_{fr} SM\_FR \\
&\quad \beta_{GA} GA,
\end{aligned}
$$

and in Figure 7 an extract from the *.mod* 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 38.

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 value of the frequency parameter β_{fr} is intuitive considering that this is actually a coefficient for headway; 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 $\mu_{classic} = 1.64$.

To be consistent with random utility theory, the inequality $\frac{\mu}{\mu_m} < 1$ with μ being normalized to 1 implies $\mu_m > 1$. Having estimated μ_m, BIOGEME performs the

```
[NLNests]
// Name paramvalue  LowerBound UpperBound  status list of alt
Classic       1.0      1   10     0         1 3
Innovative    1.0      1   10     1         2
```

Figure 7: BIOGEME snapshot

NL model					
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t-stat. 0*	Robust *t-stat. 1*
1	ASC_{car}	0.0272	0.119	0.23	
2	ASC_{sm}	0.243	0.119	2.05	
3	β_{cost}	-0.000986	0.000105	-9.36	
4	β_{car_time}	-0.00874	0.00101	-8.64	
5	β_{train_time}	-0.0113	0.000958	-11.78	
6	β_{sm_time}	-0.00995	0.00163	-6.09	
7	β_{fr}	-0.00472	0.000862	-5.48	
8	β_{ga}	5.39	0.583	9.26	
9	$\mu_{classic}$	1.64	0.132	12.42	4.86
Summary statistics					
Number of observations $= 6759$					
$\mathcal{L}(0) = -6958.42$					
$\mathcal{L}(\hat{\beta}) = -5207.79$					
$\bar{\rho}^2 = 0.250$					

Table 38: NL estimation results

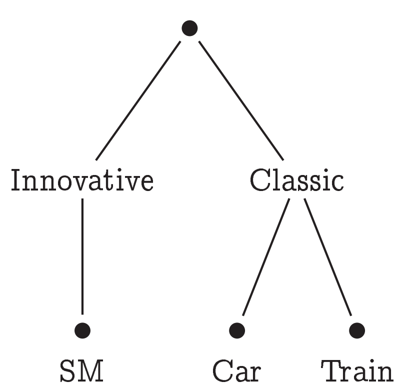

Figure 8: The correlation structure of the specified NL model

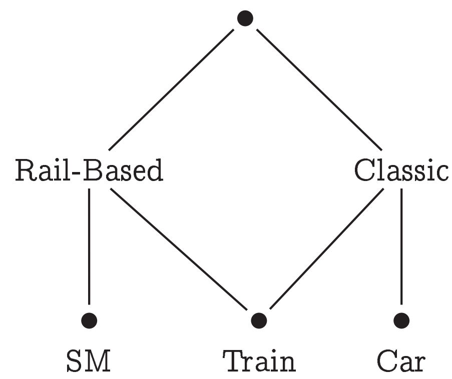

Figure 9: A representative scheme for the CNL correlation structure.

following two tests:

- First test
  Null hypothesis $H_0 : \mu_m = 0$
  Reject the null hypothesis if $|\frac{(\hat{\mu}_m - 0)}{\text{std err of } \hat{\mu}_m}| > t_{\text{critical\_value}}$
- Second test
  Null hypothesis $H_0 : \mu_m > 1$
  Reject the null hypothesis if $|\frac{(\hat{\mu}_m - 1)}{\text{std err of } \hat{\mu}_m}| > t_{\text{critical\_value}}$

Here, μ_{classic} is significantly different from one which indicates that there is a significant correlation between the car and train alternatives.

Estimation of a Cross-Nested Logit Model with Fixed Alphas

Files to use with BIOGEME:
model file: GEV_SM_CNL_fix.mod
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 9.

In Figure 10, we show a snapshot from the BIOGEME `.mod` file illustrating the CNL nest specification. The estimation results are shown in Table 39.

In this CNL specification, we have fixed the $\alpha_{\text{train\_classic}}$ and $\alpha_{\text{train\_rail}}$ coefficients to 0.5. It means that we assume that the train alternative equally belongs to both nests *classic* and *rail-based*.

The CNL specification gives us more flexibility in the definition of the correlation structure between the alternatives. In this case, where the α's are fixed, μ_{classic} and

```
[CNLNests]
// Name        paramvalue LowerBound UpperBound  status
classic           1.0        1          10        0
Rail_based        1.0        1          10        0

[CNLAlpha]
// Alt     Nest       value   LowerBound    UpperBound  status
Car      classic        1       0.00001        1.0        1
Train    classic        0.5     0.00001        1.0        1
Train    Rail_based     0.5     0.00001        1.0        1
SM       Rail_based     1       0.00001        1.0        1
```

Figure 10: BIOGEME snapshot

CNL model with fixed α's					
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t-stat. 0*	Robust *t-stat. 1*
1	ASC_{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	β_{car_time}	-0.00628	0.00122	-5.17	
5	β_{train_time}	-0.00863	0.00105	-8.18	
6	β_{sm_time}	-0.00715	0.00151	-4.74	
7	β_{fr}	-0.00298	0.000533	-5.58	
8	β_{ga}	0.618	0.0940	6.57	
9	$\mu_{classic}$	2.85	0.260	10.93	7.09
10	μ_{rail_based}	4.73	0.483	9.79	7.72
Summary statistics					
Number of observations = 6759					
$\mathcal{L}(0) = -6958.42$					
$\mathcal{L}(\hat{\beta}) = -5120.74$					
$\bar{\rho}^2 = 0.263$					

Table 39: Estimation results for the CNL specification. The α coefficients are fixed.

CNL model with unknown α's					
Variable number	Variable name	Coefficient 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_{fr}	-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	18.34	-19.40
12	α_{train_rail}	0.514	0.0265	19.40	-18.34
Summary statistics					
Number of observations = 6759					
$\mathcal{L}(0) = -6958.42$					
$\mathcal{L}(\hat{\beta}) = -5120.61$					
$\bar{\rho}^2 = 0.262$					

Table 40: Estimation results for the CNL specification. The α coefficients are estimated.

μ_{rail_based} are estimated significantly differently from one. We can thus conclude that the train alternative is correlated with both Swissmetro and car alternatives. In the next example, we are going to relax the assumption of fixed α coefficients and try to estimate them directly from the data.

Estimation of a Cross-Nested Logit Model with Unknown Alphas

Files to use with BIOGEME:
Model file: GEV_SM_CNL_var.mod
Data file: swissmetro.dat

In Table 40, we show the results for the CNL specification with unknown α coefficients. It is interesting to note that the estimated alpha parameters are not significantly different from 0.5. Fixing the alphas to 0.5 is then accepted for this model. 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.