EPFL
ENAC TRANSP-OR
**Prof. M. Bierlaire**

Mathematical Modeling of Behavior
Fall 2016

Swissmetro

# 1 Estimation of a Nested Logit Model

*Files to use with Biogeme:*
*Model file:* *GEV_SM_NL.py*
*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 both assigned to the same nest and the Swissmetro is alone in a second nest, as shown in Figure 1.
The expressions of the systematic utility functions for each alternative used in this model specification are

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

Note that only one of the two nest parameters can be estimated. The estimation results are shown in Table 1.
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 $\beta_{he}$ indicates that the higher the headway, the lower the frequency of service, and thus the lower

| NL model | | | | | |
|---|---|---|---|---|---|
| Parameter number | Parameter name | Parameter estimate | Robust standard error | Robust *t-stat. 0* | Robust *t-stat. 1* |
| 1 | $\mathrm{ASC}_{car}$ | 0.0272 | 0.119 | 0.23 | |
| 2 | $\mathrm{ASC}_{SM}$ | 0.243 | 0.119 | 2.05 | |
| 3 | $\beta_{cost}$ | -0.000986 | 0.000105 | -9.36 | |
| 4 | $\beta_{car\_time}$ | -0.00874 | 0.00101 | -8.64 | |
| 5 | $\beta_{train\_time}$ | -0.0113 | 0.000958 | -11.77 | |
| 6 | $\beta_{SM\_time}$ | -0.00995 | 0.00163 | -6.09 | |
| 7 | $\beta_{he}$ | -0.00472 | 0.000862 | -5.48 | |
| 8 | $\beta_{ga}$ | 5.39 | 0.582 | 9.26 | |
| 9 | $\mu_{classic}$ | 1.64 | 0.132 | 12.42 | 4.86 |

**Summary statistics**
Number of observations = 6759
$\mathcal{L}(0) = -6958.425$
$\mathcal{L}(\hat{\beta}) = -5207.794$
$\bar{\rho}^2 = 0.250$

Table 1: NL estimation results

the utility. Finally, the scale parameter of the random term associated with the *classic* nest has been estimated as $\mu_{classic} = 1.64$.
To be consistent with random utility theory, the inequality $\frac{\mu}{\mu_m} < 1$ with $\mu$ 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: $\frac{(\hat{\mu}_m - 1)}{\text{std err of } \hat{\mu}_m}$. It is also output by Biogeme. Here the t-statistic with respect to 1 is 4.86, 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

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-5245.512 + 5207.794) = 75.436$$

where the restricted model is the logit model (*SpecTest_SM_socioec_bis.py*) and the unrestricted model is the nested logit model. The test statistic is asymptotically $\chi^2$ distributed with 1 degree of freedom since there is 1 restriction. Since $75.436 > 3.841$ (the critical value of the $\chi^2$ 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.

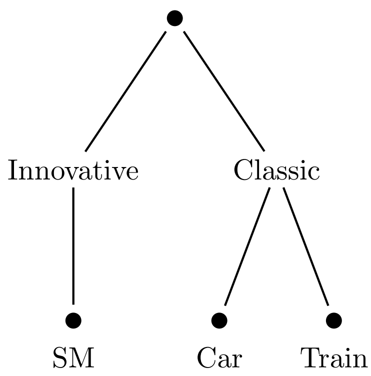


Figure 1: The correlation structure of the specified NL model

## 2 Estimation of a Cross-Nested Logit Model with Fixed Alphas

*Files to use with Biogeme:*
*model file: GEV_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 **??**.
The estimation results are shown in Table 2. The alternative-specific constants now have a negative sign. All other coefficients have the expected signs.
In this CNL specification, we have fixed the $\alpha_{train\_classic}$ and $\alpha_{train\_rail}$ 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 $\alpha$'s is a restricted model of CNL with variable $\alpha$'s.

### Estimation of a Cross-Nested Logit Model with Unknown Alphas

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

In Table 3, we show the results for the CNL specification with variable $\alpha$ coefficients. We also

| CNL model with fixed $\alpha$'s | | | | | |
|---|---|---|---|---|---|
| Parameter number | Parameter name | Parameter estimate | Robust standard error | Robust *t-stat. 0* | Robust *t-stat. 1* |
| 1 | $\mathrm{ASC}_{car}$ | -0.838 | 0.0787 | -10.65 | |
| 2 | $\mathrm{ASC}_{SM}$ | -0.457 | 0.0744 | -6.15 | |
| 3 | $\beta_{cost}$ | -0.00705 | 0.000526 | -13.39 | |
| 4 | $\beta_{car\_time}$ | -0.00628 | 0.00122 | -5.17 | |
| 5 | $\beta_{train\_time}$ | -0.00863 | 0.00105 | -8.18 | |
| 6 | $\beta_{SM\_time}$ | -0.00715 | 0.00151 | -4.74 | |
| 7 | $\beta_{he}$ | -0.00298 | 0.000533 | -5.58 | |
| 8 | $\beta_{ga}$ | 0.618 | 0.0940 | 6.57 | |
| 9 | $\mu_{classic}$ | 2.85 | 0.260 | 10.93 | 7.09 |
| 10 | $\mu_{rail\_based}$ | 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 2: Estimation results for the CNL specification. The $\alpha$ coefficients are fixed.

| CNL model with unknown $\alpha$'s | | | | | |
|---|---|---|---|---|---|
| Parameter number | Parameter name | Parameter estimate | standard error | *t-stat. 0* | *t-stat. 1* |
| 1 | $\text{ASC}_{car}$ | -0.849 | 0.0692 | -12.26 | |
| 2 | $\text{ASC}_{SM}$ | -0.460 | 0.0656 | -7.01 | |
| 3 | $\beta_{cost}$ | -0.00697 | 0.000440 | -15.85 | |
| 4 | $\beta_{car\_time}$ | -0.00621 | 0.000583 | -10.66 | |
| 5 | $\beta_{train\_time}$ | -0.00849 | 0.000660 | -12.85 | |
| 6 | $\beta_{SM\_time}$ | -0.00711 | 0.000745 | -9.54 | |
| 7 | $\beta_{he}$ | -0.00293 | 0.000510 | -5.75 | |
| 8 | $\beta_{ga}$ | 0.620 | 0.0886 | 7.00 | |
| 9 | $\mu_{classic}$ | 2.87 | 0.212 | 13.54 | 8.82 |
| 10 | $\mu_{rail\_based}$ | 4.90 | 0.722 | 6.78 | 5.40 |
| 11 | $\alpha_{train\_classic}$ | 0.486 | 0.0265 | 18.35 | -19.40 |
| 12 | $\alpha_{train\_rail}$ | 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 3: Estimation results for the CNL specification. The $\alpha$ coefficients are estimated.

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.

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

$$-2(\mathcal{L}_R - \mathcal{L}_U) = -2(-5207.794 + 5120.608) = 174.372$$

where the restricted model is the nested logit model and the unrestricted model is the CNL model with variable $\alpha$'s. The test statistic is asymptotically $\chi^2$ distributed with 2 degrees of freedom since there are 2 restrictions. Since $174.372 > 5.991$ (the critical value of the $\chi^2$ 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 $\alpha$'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 $\alpha$'s and the CNL model with variable $\alpha$'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 $\alpha_{train\_rail}$ is not significantly different from 0.5, and hence we accept the null hypothesis (CNL model with fixed $\alpha$'s) and reject the CNL model with variable $\alpha$'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 $\alpha$'s and the unrestricted model is the CNL model with variable $\alpha$'s. The test statistic is asymptotically $\chi^2$ distributed with 1 degree of freedom since there is 1 restriction. Since $0.260 < 3.841$ (the critical value of the $\chi^2$ distribution with 1 degree of freedom at a 95 % level of confidence), we accept the null hypothesis (CNL model with fixed $\alpha$'s) and reject the CNL model with variable $\alpha$'s.

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

---

mbi/ ek/ afa / mpp