

## **Specification(s) of Nested Logit Models**

Florian Heiss

**16-2002**

January 2002

# Specification(s) of Nested Logit Models\*

Florian Heiss  
University of Mannheim, MEA

January 4, 2002

## Abstract

The nested logit model has become an important tool for the empirical analysis of discrete outcomes. There is some confusion about its specification of the outcome probabilities. Two major variants show up in the literature. This paper compares both and finds that one of them (called random utility maximization nested logit, RUMNL) is preferable in most situations. Since the command `nlogit` of Stata 7.0 implements the other variant (called non-normalized nested logit, NNNL), an implementation of RUMNL called `nlogitrum` is introduced. Numerous examples support and illustrate the differences of both specifications.

## 1 Introduction

The nested logit model has become an important tool for the empirical analysis of discrete outcomes. It is attractive since it relaxes the strong assumptions of the multinomial (or conditional) logit model. At the same time, it is computationally straightforward and fast compared to the multinomial probit, mixed logit, or other even more flexible models due to the existence of a closed-form expression for the likelihood function.

---

\*I would like to thank Axel Börsch-Supan, Melanie Lührmann, and Joachim Winter for their helpful comments and the Gesamtverband der Deutschen Versicherungswirtschaft and the Land Baden Württemberg for their financial support. Correspondence to: Florian Heiss; University of Mannheim; MEA; L13,17; 68131 Mannheim; Germany. Phone: +49-621-181-1858, Fax: +49-621-181-1863, email: heiss@econ.uni-mannheim.de

There is some confusion about the specification of the outcome probabilities in nested logit models. Two substantially different formulas and many minor variations of them are presented and used in the empirical literature and in textbooks. Many researches are neither aware of this issue nor which version is actually implemented by the software they use. This obscures the interpretation of their results. This problem has been previously discussed by Hensher and Greene (2000), Hunt (2000), Koppelman and Wen (1998), and Louviere et al. (2000, section 6.5). This paper provides a comparison of both approaches in a similar spirit. It argues and shows in numerous examples that one of these specifications is preferable in most situations. The package `nlogit` of Stata 7.0 does not implement this specification. Therefore the package `nlogitrum` is presented, which does.

Section 2 provides a brief introduction to nested logit models and their different specifications. Section 3 discusses the consequences of these specifications in various modeling situations and illustrates the arguments with empirical examples. Section 4 introduces the new Stata commands and section 5 concludes.

## 2 Nested logit defined

The nested multinomial logit (NMNL) model is a generalization of the multinomial (or McFadden's conditional) logit model. It allows groups of alternatives to be similar to each other in an unobserved way. The researcher specifies a structure that partitions the alternatives into groups ('nests'). This can be generalized to various nesting levels by grouping the alternatives within such a nest in sub-nests and so on.

For simplicity, assume a two-level nesting structure. Suppose that there are  $J$  alternatives numbered as  $j = 1, 2, \dots, J$ . They are partitioned into  $K$  nests  $N_1, N_2, \dots, N_k, \dots, N_K$ . Suppose,  $y \in \{1, 2, \dots, j, \dots, J\}$  is an indicator for the realized outcome. If alternative  $j$  is an element of nest  $N_k$ , then the probability of  $y = j$  can in general be decomposed into

$$P(y = j) = P(y \in N_k) \cdot P(y = j | y \in N_k). \quad (1)$$

The NMNL model can be derived from a structural model of random utility maximization (RUM). Agents are assumed to choose the alternative from which they derive the highest utility. The utility for alternative  $j$  is

specified as

$$U_j = \mathbf{x}'_j \boldsymbol{\beta} + \mathbf{z}' \boldsymbol{\gamma}_j + \varepsilon_j. \quad (2)$$

It is determined by a set of individual-specific (socio-economic) explanatory variables  $\mathbf{z}$  containing a constant, by a set of alternative-specific variables  $\mathbf{x}_j$ , and by a random error term  $\varepsilon_j$ . If these error terms are assumed to be distributed according to a special form of the generalized extreme value (GEV) distribution, the resulting outcome probabilities are:<sup>1</sup>

$$P(y = j | y \in N_k) = \frac{e^{\frac{1}{\tau_k} V_j}}{e^{IV_k}}, \quad (3)$$

where  $V_j = \mathbf{x}'_j \boldsymbol{\beta} + \mathbf{z}' \boldsymbol{\gamma}_j$  denotes the deterministic part of the utility, and

$$P(y \in N_k) = \frac{e^{\tau_k IV_k}}{\sum_m e^{\tau_m IV_m}}, \quad (4)$$

with the inclusive value  $IV_k$  defined as

$$IV_k = \ln \sum_{l \in N_k} e^{\frac{1}{\tau_k} V_l}. \quad (5)$$

The marginal probability of the outcome  $j$  according to equation (1) is under these assumptions equal to

$$P(y = j) = \frac{e^{\frac{1}{\tau_k} V_j}}{e^{IV_j}} \cdot \frac{e^{\tau_k IV_k}}{\sum_m e^{\tau_m IV_m}}. \quad (6)$$

The parameters  $\tau_k$  are called IV or dissimilarity parameters. They correspond to the degree of dissimilarity between the alternatives within one nest.<sup>2</sup> The multinomial (conditional) logit model follows in the special case of  $\tau_k = 1, \forall k = 1, \dots, K$ . The model is consistent with RUM if all  $\tau$  lie in the unit interval.<sup>3</sup> Therefore, this model will be called random utility maximization nested logit (RUMNL) in this paper. For an introduction to this model see also Train (1986).

---

<sup>1</sup>See McFadden (1981) or Börsch-Supan (1987).

<sup>2</sup>Other equivalent parameterizations are used in the literature. McFadden (1981) writes the equivalent of equation (6) and replaces  $\tau_k$  with  $\sigma_k = 1 - \tau_k$ , Louviere et al. (2000) replaces  $\tau_k$  with  $\mu_k = 1/\tau_k$ .

<sup>3</sup>This condition can be relaxed for *local* consistency with RUM, see Börsch-Supan (1990).

If there are socioeconomic variables whose coefficients are assumed to be constant within the nests, they can instead of including them in the vector  $\mathbf{z}$  be collected in a vector  $\mathbf{w}$ . Then, an equivalent formulation of equation (6) is

$$P(y = j) = \frac{e^{\frac{1}{\tau_k} V_j}}{e^{IV_j}} \cdot \frac{e^{\mathbf{w}'\boldsymbol{\delta}_k + \tau_k IV_k}}{\sum_m e^{\mathbf{w}'\boldsymbol{\delta}_m + \tau_m IV_m}}. \quad (7)$$

A similar model that is also called nested logit is presented in the literature, see for example Greene (2000, section 19.7.4). The choice probabilities of this model are specified as

$$Q(y = j) = \frac{e^{V_j}}{e^{IV_j}} \cdot \frac{e^{\mathbf{w}'\boldsymbol{\delta}_k + \theta_k IV_k}}{\sum_m e^{\mathbf{w}'\boldsymbol{\delta}_m + \tau_m IV_m}}, \quad (8)$$

with

$$IV_k = \ln \sum_{l \in N_k} e^{V_l}. \quad (9)$$

This is the model implemented by `nlogit`, see [R] `nlogit`. The relationship between both specifications becomes more evident if the RUMNL probability (7) is rewritten as

$$P(y = j) = \frac{e^{\tilde{V}_j}}{e^{\tilde{IV}_j}} \cdot \frac{e^{\mathbf{w}'\boldsymbol{\delta}_k + \tau_k \tilde{IV}_k}}{\sum_m e^{\mathbf{w}'\boldsymbol{\delta}_m + \tau_m \tilde{IV}_m}}, \quad (10)$$

with

$$\tilde{IV}_k = \ln \sum_{l \in N_k} e^{\tilde{V}_l}. \quad (11)$$

and

$$\tilde{V}_j = \frac{V_j}{\tau_k}. \quad (12)$$

A comparison with equations (8) and (9) shows that this model corresponds to a RUMNL model except that the  $V_j$  are replaced the the scaled  $\tilde{V}_j$  with different scaling factors  $1/\tau_k$  across nests. Therefore, this model will be called non-normalized nested logit (NNNL) model in this paper.

Because of the scaling, the parameter estimates of NNNL are not direct estimates of the parameters of the underlying RUM model. Depending on the specification of  $V_j$  and the nesting structure, this can have different implications for the consistency of the model with the underlying theory as will be discussed in detail in the next section.

### 3 Different model specifications

In this section, the differences between RUMNL and NNNL in different model settings are discussed and demonstrated in various examples. The empirical examples are based on well-known data on travel mode choice. Among others, Greene (2000, example 19.18), Hunt (2000), and Louviere et al. (2000, section 6.4) present nested logit estimates based on them. The data contain 210 non-business travelers between Sydney, Canberra, and Melbourne. They had four travel modes alternatives: car, train, bus, plane. In this paper, only three sets of explanatory variables are used to demonstrate the effects of the different model specifications. These are:

- **const** × *alt*: alternative-specific constants (ASCs)
- **hinc** × *alt*: Household income, interacted with the ASCs
- **time** and **time** × *alt*: traveling time with the alternative modes, (partly) interacted with the ASC

#### 3.1 The case of alternative-specific coefficients only

If only alternative-specific coefficients  $\gamma_j$  enter the model, equation (2) simplifies to

$$V_j = \mathbf{z}'\gamma_j \quad (13)$$

and equation (12) can be written as

$$\tilde{V}_j = \mathbf{z}'\tilde{\gamma}_j \quad (14)$$

with

$$\tilde{\gamma}_j = \frac{1}{\tau_m}\gamma_j. \quad (15)$$

The parameters  $\tilde{\gamma}_j$  of a NNNL model do not directly have a structural interpretation in a RUM model. The estimates of the underlying parameters  $\gamma_j$  can be recovered after estimation as  $\hat{\gamma}_j = \hat{\tilde{\gamma}}_j \hat{\tau}_k$ , where  $\hat{\tilde{\gamma}}_j$  and  $\hat{\tau}_k$  are the parameter estimates of  $\tilde{\gamma}_j$  and  $\tau_k$ , respectively. This should be kept in mind when the parameters are interpreted.

In order to illustrate this point, table 1 shows the results for both a RUMNL (Model A) and a NNNL (Model B). The chosen nesting structure is depicted in figure 1: The public transportation modes (train and bus)

as well as the other modes (air and car) share a nest. Both models are equivalent in terms of the log likelihood. The estimated IV parameters are also identical. But the other parameter estimates differ. As discussed above, the RUMNL model directly estimates the structural parameters  $\gamma_j$ . They can be recovered from the NNNL estimates by rescaling them. For example, the estimate for the structural parameter of ‘hinc×train’ is  $\hat{\gamma}_{\text{hinc}\times\text{train}} = -0.047$ . It can be recovered from the NNNL estimates by multiplying its coefficient  $\tilde{\gamma}_{\text{hinc}\times\text{train}} = -0.088$  with the estimated IV parameter of the respective nest  $\hat{\tau}_{\text{public}} = 0.539$  as can be easily verified:  $-0.047 = -0.088 \times 0.539$ .

This scaling impedes the direct interpretation of the NNNL estimates and of tests based on them. For example the presented asymptotic t-statistic for ‘hinc×train’ for the NNNL model does not test whether income has an effect on the probability of taking the train relative to taking the car. The appropriate null hypothesis is  $H_0 : \tilde{\gamma}_{\text{hinc}\times\text{train}} \times \tau_{\text{public}} = 0$ . Likewise, the test of the null hypothesis  $H_0 : \gamma_{\text{time}\times\text{air}} = \gamma_{\text{time}\times\text{bus}}$  for the RUMNL model is equivalent to the test of null hypothesis  $H_0 : \tilde{\tau}_{\text{time}\times\text{air}} \times \tau_{\text{other}} = \tilde{\gamma}_{\text{time}\times\text{bus}} \times \tau_{\text{public}}$  for the NNNL model. For the RUMNL model, these hypotheses can be tested in a standard way:<sup>4</sup>

```
. nlogitrum mode asc_* hinc_* time_*, group(grp) nests(travel type)
(output omitted)
. test hinc_train = 0
( 1) [travel]hinc_train = 0.0
      chi2( 1) =    6.07
      Prob > chi2 =  0.0138
. test time_air=time_bus
( 1) [travel]time_air - [travel]time_bus = 0.0
      chi2( 1) =   26.45
      Prob > chi2 =  0.0000
```

These direct tests are invalid after NNNL estimation. It would be wrong to test these hypotheses after the NNNL model as

```
. nlogit mode (travel = asc_* hinc_* time_*)(type=temp), group(grp) const (1)
(output omitted)
. test hinc_train = 0
( 1) [travel]hinc_train = 0.0
      chi2( 1) =    4.02
      Prob > chi2 =  0.0450
. test time_bus=time_air
( 1) - [travel]time_air + [travel]time_bus = 0.0
```

---

<sup>4</sup>The syntax of `nlogitrum` will be introduced in section 4.

```

chi2( 1) = 2.23
Prob > chi2 = 0.1352

```

Instead, the correct tests that are equivalent to those after the RUMNL estimation can be obtained from the NNNL models if the scaling is taken into account:

```

. testnl _b[hinc_train]*[public]_cons = 0
(1)  _b[hinc_train]*[public]_cons = 0
      chi2(1) = 6.07
      Prob > chi2 = 0.0138
. testnl _b[time_air]*[other]_cons = _b[time_bus]*[public]_cons
(1)  _b[time_air]*[other]_cons = _b[time_bus]*[public]_cons
      chi2(1) = 26.45
      Prob > chi2 = 0.0000

```

So in the case of alternative-specific coefficients, both models are equivalent. But while the RUMNL model directly estimates the parameters of interest, the estimated coefficients from NNNL have to be rescaled before they are to be interpreted and when tests are based on them. For example, the asymptotic t-statistics from the output of `nlogit` do not correspond to tests of intrinsically interesting hypotheses.

### 3.2 The case of generic variables

The alternative-specific utility can also be affected by attributes of the alternative such as the price of the transportation mode. This factor can be allowed for in RUM models by introducing a single parameter for each of these variables that are called generic variables. In equation (2), this is done by including  $\mathbf{x}'_j\boldsymbol{\beta}$  in the specification of the utility: The subscript for the alternative belongs to the attributes  $\mathbf{x}_j$  and the parameters  $\boldsymbol{\beta}$  are constant across alternatives.

If generic variables are included in the model, the NNNL specification is inconsistent with the underlying RUM model. Analogous to the case of alternative-specific coefficients, the NNNL scales the coefficients of generic variables with a scaling factor that varies over the nests. But unlike the former case, the RUM model requires that the parameters  $\boldsymbol{\beta}$  are constant across alternatives. NNNL replaces this restriction with the restriction of constant *scaled* parameters  $\frac{1}{\tau_m}\boldsymbol{\beta}$ . This makes the econometric model inconsistent with the underlying RUM model as long as the dissimilarity parameters vary across nests.



The differences between RUMNL and NNNL for the case of generic variables are demonstrated by the estimates shown in table 2. The coefficients for the variable ‘time’ in table 1 are very similar for the alternatives ‘car’, ‘train’, and ‘bus’ and the hypothesis that they are actually equal can not be rejected. So table 2 shows the results for a modified specification in which ‘time’ enters as a generic variable. Since the ‘time’ coefficient for ‘air’ significantly differs from the others in the previous estimates, an interaction term for ‘time’ with ‘air’ is added. This modification corresponds to constraining the other three ‘time’ coefficients to be equal to each other.

A comparison of models C and D in table 2 shows that the NNNL model does not give the same estimates as the RUMNL model in this case. In particular, the log likelihood differs between both models.<sup>5</sup> This is because the IV parameters not only constitute the (dis)similarities, but also the different scaling of the generic variable across nests. As a result, model D is misspecified.

If one is willing to assume a priori that the dissimilarity parameters of all nests in a nesting level have the same value, the problem of a misspecified NNNL model disappears. This is demonstrated in models E and F in table 2. In this case, the results from RUMNL and NNNL only differ in the scaling of the parameters. The presence of the generic variable does not distort the estimates of the NNNL model, since its parameter is forced to be scaled equally in each nest. The problem with this constraint is that it cannot be tested with NNNL estimates, because the unconstrained model D is misspecified. In contrast, both RUMNL specifications are valid and a comparison of the log likelihood values from models C and E clearly shows that this constraint is rejected by the data.

### 3.3 Degenerate nests

If a nest only contains one alternative, it is called degenerate nest. The dissimilarity parameter of degenerate nests is not identified in the RUMNL model. This can be easily seen from equations (3) to (6). Since nest  $N_k$  only contains alternative  $j$ , its inclusive value simplifies to  $IV_k = \frac{1}{\tau_k} V_j$ . The conditional probability  $P(y = j | y \in N_k)$  is trivially equal to unity. The

---

<sup>5</sup>The point here is not that the likelihood value of the NNNL model is *lower*. That does not need to be the case if the restrictions imposed by the specification of a generic variable does not hold. It is important to note that it is *different*.

dissimilarity parameter  $\tau_k$  cancels out in the probability of choosing nest  $N_k$ , so it appears neither in the marginal probability of choosing alternative  $j$ , nor in the likelihood function. This is intuitive since the concept of (dis)similarity does not make sense with only one alternative.

In the NNNL model however, the dissimilarity parameter does not vanish from the choice probability and may be statistically identified. If only alternative-specific parameters  $\tilde{\gamma}_j$  are included in the model, they are not jointly identified with the dissimilarity parameter in degenerate nests. So the dissimilarity parameter can be constrained to any nonzero value. The only effect of choosing this value is that the ML estimates of  $\tilde{\gamma}_j$  are scaled accordingly.

If at least one generic variable is included in the NNNL model, the IV parameter of degenerate nests may be identified along with the other model parameters. This identification comes from the restriction of equal *scaled* parameters  $\frac{1}{\tau_m}\beta$  across alternatives and nests and the parameters only constitute this scaling. A conventional approach to restrict the IV parameter to be equal to unity does not result in a model that is consistent with the underlying RUM model.

This is demonstrated with the estimates shown in table 3. The fact that the estimated dissimilarity parameter of the nest ‘other’ in table 2 is substantially larger than 1 indicates that the alternatives ‘air’ and ‘car’ should not share a nest. Therefore, the nesting structure is modified by splitting this nest into two degenerate nests. The resulting nesting structure is depicted in figure 2. In models G and H shown in table 3, the variable ‘time’ purely enters as a generic variable. The dissimilarity parameters of the degenerate nests ‘air’ and ‘car’ are not identified from the RUMNL model G. As argued above, they cancel out of the likelihood function. In contrast, all IV parameters are identified in the NNNL model H. It has two more free parameters than the RUMNL model and a substantially higher likelihood value.

However, these IV parameters do not have anything to do with (dis)similarity. They simply relax the constraint of equal scaling of the generic variable coefficient across nests. To demonstrate this, models I and J shown in table 3 do the same explicitly by estimating a separate ‘time’ coefficient for each nest. As a result, the IV parameters of the degenerate nests are not jointly identified with the other parameters of the corresponding nests in the NNNL model and have to be constrained to any nonzero number. Both models result in the same log likelihood value and the parameters are equivalent if the NNNL parameters are rescaled with the value of the corresponding IV

parameter. The results are also equivalent to model H. This supports the assertion that the IV parameters in model H do nothing more than relax the constraint of equal scaling.

So if there is only one generic variable present in the model, the NNNL estimate of the IV parameter can be interpreted in a straightforward way, although this is probably not the way the researcher intends to interpret IV parameters. It is much more direct to explicitly relax the specification of generic variables. If there are more than one generic variables, the interpretation becomes more obscure. Then the NNNL specification imposes the restriction that the coefficients of all generic variables differ proportionally across nests. Greene (2000, example 19.18) presents model that has this problem. It is a NNNL model based on the same data used in this paper. In addition to ‘time’, the generic variable ‘cost’ is included. As a result, the estimates have no clear interpretation. The RUMNL avoids the danger of misspecification and misinterpretation.

### 3.4 Dummy nests

There is a way to ‘trick’ NNNL software into estimating a RUM consistent nested logit model independent of the type of included explanatory variables and without imposing equality of dissimilarity parameters. Koppelman and Wen (1998) propose to add degenerate dummy nests and constrain their IV parameters appropriately. This can most easily explained by an example.

Figure 3 shows the nesting structure for the travel mode choice example according to figure 1 with appropriate dummy nests added. For each alternative, such a degenerate nest is specified. The corresponding IV parameters  $\rho_1$  through  $\rho_4$  are shown next to each nest along with the respective constraint. The two ‘Public’ alternatives each have a degenerate dummy nest whose IV parameters are constrained to be equal to the IV parameter of the ‘Other’ nest. Intuitively, their parameters are first scaled by  $1/\tau_1$ . Then the additional dummy nest scales them by  $1/\tau_2$ . For the two ‘Other’ alternatives, this works accordingly. As a result, the parameters of *all* alternatives are scaled by  $\frac{1}{\tau_1\tau_2}$ . While  $\tau_1$  and  $\tau_2$  can be allowed to differ, this does not translate into different scaling across nests.

Table 4 shows the results from a specification according to this strategy. Model K shown there is identical to model C shown in table 2. It could not be reproduced by NNNL since it contains generic variables and the IV parameters are allowed to differ between nests. Model L is a NNNL model with

the dummy nests added as described above. As can be seen, this specification mimics the RUMNL model except for the scaling of the parameters for the explanatory variables. The structural coefficients can be recovered from these estimates by multiplying the estimated coefficients by both estimated IV parameters. For example, the coefficient for 'hinc×train is -0.047. it can be calculated from the NNNL-DN estimates as  $-0.047 = -0.018 \times 4.801 \times 0.545$ .

Depending on the original nesting structure, a large number of dummy nests may be needed for this strategy. This complicates both the specification and the estimation.<sup>6</sup> This strategy therefore seems to be a real alternative to RUMNL for researchers who only have access to a NNNL implementation.

## 4 Stata implementation

The NNNL model is available for Stata 7.0 users as the `nlogit` command. As argued in this paper, the RUMNL model is preferable in most situations. This section introduces the command `nlogitrum.ado` that implements the RUMNL model and that was used to produce all RUMNL estimates in this paper. Furthermore, the command `nlogitdn.ado` is described. It adds dummy nests to any specified nesting structure as discussed in section 3.4.

### 4.1 Data setup

The data setup for `nlogitrum` is equivalent to `nlogit`. That is, a set of categorical variables `altsetvarB [ ... altsetvar2 altsetvar1]` is generated using `nlogitgen`. The tree structure can be visualized using `nlogittree`. For a thorough description see [R] `nlogit`.

### 4.2 Syntax

```
nlogitrum depvar indepvars [weight] [if exp] [in range] , group(varname)
      nests(altsetvarB [ ... altsetvar2 altsetvar1]) [ notree nolabel clogit
      level(#) nolog robust ivconstraints(string) constraints(numlist)
      maximize_options ]
```

---

<sup>6</sup>The command `nlogitdn`, introduced in section 4.4, automates the generation of dummy nests and appropriate constraints.

The syntax is similar to that of `nlogit` with one major difference. `nlogit` insists on having explanatory variables for each nesting level and `nlogitrum` only allows explanatory variables to directly enter the conditional probabilities of the alternatives. There are three reasons for this change. The first reason is that in many cases the user cannot think of a variable that is specific to a nest instead of an alternative. So one often ends up throwing nonsense variables into the specification of nest-specific explanatory variables and constraining their coefficients to zero. The second reason is that for the RUMNL model, it does not make a difference at all if a nest-specific variable is specified for a nest or for all alternatives within the nest, see equation (7). The third reason is that it greatly simplifies the syntax and makes it equivalent to the syntax of `cllogit` except for the additional options.

The option `d1` of `nlogit` does not exist for `nlogitrum`. The current version uses the `m1` method `d0`.

### 4.3 Predictions

The syntax for `predict` after `nlogitrum` is almost identical to the syntax after `nlogit` estimation. The only difference is that the options `xbb` and `xbb#` are replaced by the option `xb`, since the linear prediction can only be sensibly defined for the bottom level (the alternatives).

### 4.4 Generating dummy nests: `nlogitdn`

The command `nlogitdn` is a wrapper for `nlogit`. Its syntax is equivalent to the `nlogit` syntax. `nlogitdn` analyzes the specified nesting structure, adds appropriate dummy nests and constraints to the specification as discussed in section 3.4, and calls `nlogit`. It was used for the estimation of model L in table 4.

### 4.5 Examples

In order to help the reader to become accustomed to the syntax, the commands used to produce the example models A through L are listed below. Most variable names should be self-explanatory. The variable `grp` identifies the observations and the variable `travel` identifies the alternatives and takes the values 0 for air, 1 for train, 2 for bus, and 3 for car. The variable `mode` is the 0/1 coded dependent variable. For most models, the nesting structure

is depicted in figure 1. The respective variable `type` was generated using `nlogitgen`. For the models G through J, the nesting structure according to figure 1 was generated with the variable `typedeg`:

```
. nlogitgen type = travel(public: 1 | 2, other: 0 | 3 )
new variable type is generated with 2 groups
lb_type:
    1 public
    2 other

. nlogitgen typedeg = travel(public: 1 | 2, air: 0, car: 3)
new variable typedeg is generated with 3 groups
lb_typedeg:
    1 public
    2 air
    3 car
```

Since no variables enter the models on the level of the nests, the nonsense variables `nothing1` and `nothing2` were generated. The constraints that show up in the `nlogit` commands constrain their coefficients to zero. The models themselves were estimated using the following commands:

```
. * Model A:
. nlogitrum mode asc_* hinc_* time_*, group(grp) nests(travel type)
(output omitted)

. * Model B:
. nlogit mode (travel = asc_* hinc_* time_* )(type=nothing1), group(grp) const
> (1)

. * Model C:
. nlogitrum mode asc_* hinc_* time time_air, group(grp) nests(travel type)
(output omitted)

. * Model D:
. nlogit mode (travel = asc_* hinc_* time time_air)(type=nothing1), group(grp)
> const (1)
(output omitted)

. * Model E:
. nlogitrum mode asc_* hinc_* time time_air, group(grp) nests(travel type) ivc(
> other=public)
(output omitted)

. * Model F:
. nlogit mode (travel = asc_* hinc_* time time_air)(type=nothing1), group(grp)
> const(1) ivc(other=public)
(output omitted)

. * Model G:
. nlogitrum mode asc_* hinc_* time , group(grp) nests(travel typedeg) ivc(air=3
> .14159, car=3.14159)
(output omitted)

. * Model H:
. nlogit mode (travel = asc_* hinc_* time)(typedeg=nothing2), group(grp) const(
> 2)
(output omitted)

. * Model I:
. nlogitrum mode asc_* hinc_* timepublic time_air time_car, group(grp) nests(tr
```

```

> avel typedeg) ivc(air=3.14159, car=3.14159)
(output omitted)
. * Model J:
. nlogit mode (travel = asc_* hinc_* timepublic time_air time_car)(typedeg=noth
> ing2), group(grp) const(2) ivc(air=1, car=1)
(output omitted)
. * Model K = Model C
. * Model L:
. nlogitdn mode (travel = asc_* hinc_* time time_air)(type=nothing1), group(grp
> ) const (1)
(output omitted)

```

Note that the IV parameters of ‘air’ and ‘car’ in models G and I do not actually exist as discussed in section 3.3. Since the algorithm does not realize this beforehand, these parameters have to be restricted to an arbitrary nonzero number (in the examples, 3.14159 was chosen).

## 5 Conclusions

The name ‘nested logit’ has been given to different models. This paper argues and demonstrates that the seemingly slight difference in the specification of the outcome probabilities can lead to substantially different results and interpretations thereof. So researchers using a nested logit model (and the readers of their results) should be aware of the actual variant used.

One of these variants (called RUMNL in this paper) can be derived from theoretical models such as random utility maximization models that are widely used by econometricians. Therefore, the estimated coefficients can be readily interpreted and simple tests like an asymptotic t-test directly test hypotheses of interest. This holds irrespective of the type of included explanatory variables and specified nesting structure.

The alternative (called NNNL in this paper) is implemented in the package `nlogit` in Stata 7.0. Depending on the model specification, it can give equivalent results to those of RUMNL and the structural parameters can be recovered. But in order to do so, the estimated coefficients have to be rescaled and this also has to be kept in mind for hypothesis tests. This is the case if only alternative-specific parameters enter the model. If generic variables (variables with a common coefficient across alternatives) enter the model, the NNNL model places restrictions on the parameters that are often counterintuitive. The reason for this is that the inclusive value parameters in this case not only constitute the (dis)similarities of the alternatives, but

also the different scaling of the generic variable coefficients across nests. Examples for numerous important cases support the the produced arguments and help to understand the effects.

Finally, this paper introduces the Stata package `nlogitrum` that implements the preferred RUMNL model.

## References

- Börsch-Supan, A. 1987. *Econometric Analysis of Discrete Choice*. Berlin, Heidelberg: Springer.
- . 1990. On the Compatibility of Nested Logit Models with Utility Maximization. *Journal of Econometrics* 43: 373–388.
- Greene, W. H. 2000. *Econometric Analysis*. 4th ed. London: Prentice Hall.
- Hensher, D. A. and W. H. Greene. 2000. Specification and Estimation of the Nested Logit Model: Alternative Normalizations. Mimeo, New York University.
- Hunt, G. L. 2000. Alternative Nested Logit Model Structures and the Special Case of Partial Degeneracy. *Journal of Regional Science* 40: 89–113.
- Koppelman, F. S. and C.-H. Wen. 1998. Alternative Nested Logit Models: Structure, Properties and Estimation. *Transportation Research-B* 32: 289–298.
- Louviere, J. J., D. A. Hensher, and J. D. Swait. 2000. *Stated Choice Methods*. Cambridge: Cambridge University Press.
- McFadden, D. 1981. Econometric Models of Probabilistic Choice. In *Structural Analysis of Discrete Data and Econometric Applications*, eds. C. F. Manski and D. L. McFadden, 198–272. Cambridge, MA: MIT Press.
- Train, K. 1986. *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.



Figure 1: Nesting structure for models A through F

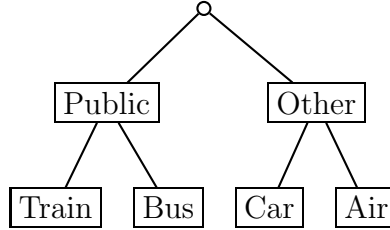


Table 1: Alternative-specific coefficients only

Model	(A)		(B)	
	RUMNL		NNNL	
	Coef.	z	Coef.	z
const × air	5.751	1.60	1.179	1.29
train	4.498	4.19	8.342	3.07
bus	3.252	2.82	6.030	2.40
hinc × air	0.035	0.90	0.007	0.90
train	-0.047	-2.46	-0.088	-2.00
bus	-0.020	-1.01	-0.038	-0.95
time × air	-0.117	-5.49	-0.024	-3.63
train	-0.022	-5.54	-0.040	-4.87
bus	-0.021	-5.37	-0.040	-4.92
car	-0.022	-5.12	-0.005	-5.03
$\tau$ other	4.879	3.58	4.879	3.58
$\tau$ public	0.539	3.69	0.539	3.69
Log likelihood	-165.12		-165.12	

Table 2: Generic variables

Constraints Model	Unconstrained IV parameters				Constrained IV parameters			
	(C)		(D)		(E)		(F)	
	RUMNL		NNNL		RUMNL		NNNL	
	Coef.	z	Coef.	z	Coef.	z	Coef.	z
const × air	6.383	2.24	2.325	2.56	6.645	3.27	2.556	3.01
train	4.597	4.90	1.101	3.13	3.114	4.40	1.198	3.64
bus	3.601	3.88	0.056	0.13	0.410	0.40	0.158	0.39
hinc × air	0.036	0.93	0.014	1.34	0.039	1.47	0.015	1.47
train	-0.047	-2.48	-0.021	-2.39	-0.052	-2.68	-0.020	-2.38
bus	-0.019	-0.98	-0.005	-0.51	-0.011	-0.42	-0.004	-0.41
time	-0.022	-5.60	-0.008	-6.75	-0.020	-5.67	-0.008	-6.73
time × air	-0.098	-5.54	-0.033	-5.39	-0.090	-5.49	-0.035	-6.04
$\tau$ other	4.801	3.84	2.638	4.36	2.600	4.40	2.600	4.40
$\tau$ public	0.545	3.79	2.535	4.29	2.600	4.40	2.600	4.40
Log likelihood	-165.26		-194.01		-194.29		-194.29	

Figure 2: Nesting structure for models G through J

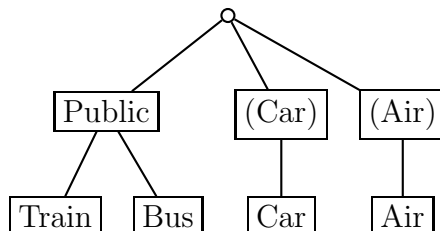


Table 3: Degenerate Nests

Model	(G)		(H)		(I)		(J)	
	RUMNL		NNNL		RUMNL		NNNL	
	Coef.	z	Coef.	z	Coef.	z	Coef.	z
const × air	-1.140	-1.97	3.158	3.41	3.613	3.83	3.613	3.83
train	2.231	5.58	13.231	3.14	2.604	4.35	13.231	3.14
bus	2.066	5.14	11.078	2.79	2.180	3.69	11.078	2.79
hinc × air	0.001	0.10	0.011	1.09	0.013	1.09	0.013	1.09
train	-0.049	-4.79	-0.236	-2.74	-0.046	-4.25	-0.236	-2.74
bus	-0.044	-4.20	-0.167	-2.19	-0.033	-2.94	-0.167	-2.19
time	-0.003	-3.79	-0.039	-4.66				
time × public					-0.008	-6.17	-0.039	-4.66
air					-0.044	-6.73	-0.044	-6.73
car					-0.007	-6.11	-0.007	-6.11
$\tau$ public	0.073	2.96	0.197	3.78	0.197	3.78	0.197	3.78
$\tau$ air	—	—*	1.144	3.86	—	—*	1	—**
$\tau$ car	—	—*	0.186	3.74	—	—*	1	—**
Log likelihood	-212.45		-182.57		-182.57		-182.57	

\*: Parameter does not exist.

\*\* : Parameter normalized to 1.

Figure 3: Nesting structure with dummy nests

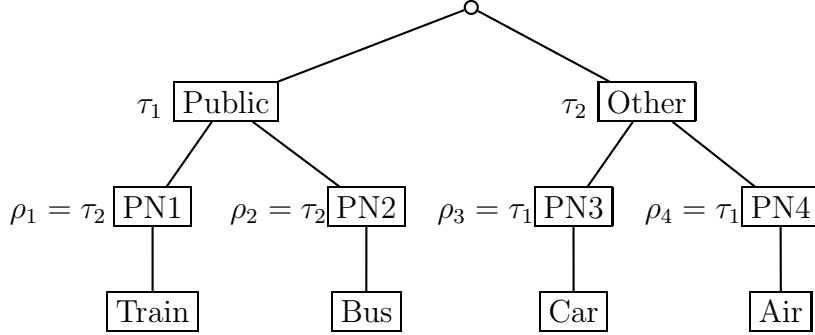


Table 4: Dummy nests

Model	(K)=(C)		(L)	
	RUMNL		NNNL-DN	
	Coef.	z	Coef.	z
const $\times$ air	6.383	2.24	2.438	1.51
train	4.597	4.90	1.756	2.61
bus	3.601	3.88	1.375	2.40
hinc $\times$ air	0.036	0.93	0.014	0.92
train	-0.047	-2.48	-0.018	-1.63
bus	-0.019	-0.98	-0.007	-0.86
time	-0.022	-5.60	-0.008	-3.08
time $\times$ air	-0.098	-5.54	-0.037	-2.69
$\tau$ other	4.801	3.84	4.801	3.84
$\tau$ public	0.545	3.79	0.545	3.79
Log likelihood	-165.26		-165.26	

## Discussion Paper Series

Mannheim Research Institute for the Economics of Aging Universität Mannheim

To order copies, please direct your request to the author of the title in question.

Nr.	Autoren	Titel	Jahr
01-02	Axel Börsch-Supan	Eine Blaupause für eine nachhaltige Rentenreform in Deutschland	02
02-02	Axel Börsch-Supan	A Blue Print For Germany's Pension Reform	02
03-02	Axel Börsch-Supan	Reduction of Working Time: Does it Decrease Unemployment?	02
04-02	Axel Börsch-Supan	Übergang statt Untergang: Wider die Furcht vor dem Untergang der Arbeit	02
05-02	Daniel Houser, Joachim Winter	How Do Behavioral Assumptions Affect Structural Inference? Evidence Form A Laboratory Experiment	02
06-02	Axel Börsch-Supan	International Comparison of Household Savings Behaviour: The German Savings Puzzle	02
07-02	Tito Boeri, Axel Börsch-Supan, Guido Tabellini	Would you Like to Reform the Pension System? The Opinions of European Citizens	02
08-02	Axel Börsch-Supan, Florian Heiss, Miki Seko	Housing Demand in Germany and Japan Paper in memoriam of Stephen King	02
09-02	Moshe Ben-Akiva, Daniel McFadden, Kenneth Train, Axel Börsch-Supan, et.al.	Hybrid Choice Models: Progress and Challenges	02
10-02	Axel Börsch-Supan, Alexander Ludwig, Joachim Winter	Aging and International Capital Flows	02
11-02	Axel Börsch-Supan	Labor market effects of population aging	02
12-02	Axel Börsch-Supan	Kann die Finanz- und Sozialpolitik die Auswirkungen der Bevölkerungsalterung auf den Arbeitsmarkt lindern?	02
13-02	Florian Heiß, Jens Köke	Dynamics in ownership and firm survival: Evidence from corporate Germany	02
14-02	Axel Börsch-Supan	„Global Aging“ an der Jahrtausendwende: Die demographischen Herausforderungen des 21. Jahrhunderts	02
15-02	Axel Börsch-Supan	Nach der Reform ist vor der Reform: Weitere Schritte für eine nachhaltige Reform der Altersvorsorge in Deutschland	02
16-02	Florian Heiss	Specification(s) of Nested Logit Models	02

17-02	Axel Börsch-Supan	What We Know and What We Do NOT Know About the Willingness to Provide Self-Financed Old-Age Insurance	02
18-02	Axel Börsch-Supan, Agar Brugiavini	Savings: The Policy Debate in Europe	02