ITS

Specification and Estimation of Nested Logit Models

by

David A. Hensher &
William H. Greene

February, 1999

*Established and supported under the Australian Research
Council's Key Centre Program.*

# INSTITUTE OF TRANSPORT STUDIES

The Australian Key Centre
in Transport Management

The University of Sydney
and Monash University

| | |
|---|---|
| **NUMBER:** | Working Paper ITS-WP-99-4 |
| **TITLE:** | Specification and Estimation of Nested Logit Models |

**ABSTRACT:** The nested logit model is currently the preferred extension to the simple multinomial logit discrete choice model. The appeal of the nested logit model is its ability to accommodate differential degrees of interdependence (i.e. similarity) between subsets of alternatives in a choice set. The received literature displays a frequent lack of attention to the very precise form that a nested logit model must take to ensure that the resulting model is invariant to normalisation of scale and is consistent with utility maximisation. Some recent papers by Koppelman and Wen (1998a, 1998b) and Hunt (1998) have addressed some aspects of this issue, but some important points remain somewhat ambiguous.

When utility function parameters have different implicit scales, imposing equality restrictions on common attributes associated with different alternatives (i.e. making them generic) can distort these differences in scale. Model scale parameters are then 'forced' to take up the real differences that should be handled via the utility function parameters. With many variations in model specification appearing in the literature, comparisons become difficult, if not impossible, without clear statements of the precise form of the nested logit model. There are a number of approaches to achieving this, with some or all of them available as options in commercially available software packages. This note seeks to clarify the issue, and to establish the points of similarity and dissimilarity of the different formulations that appear in the literature.

**AUTHORS:** David A. Hensher &
William H. Greene[**]

**CONTACT:** Institute of Transport Studies (Sydney & Monash)
The Australian Key Centre in Transport Management
C37, The University of Sydney  NSW  2006
Australia

| | |
|---|---|
| Telephone: | +61 2 9351 0071 |
| Facsimile: | +61 2 9351 0088 |
| E-mail: | itsinfo@its.usyd.edu.au |
| Internet: | http://www.its.usyd.edu.au |

| | |
|---|---|
| **DATE:** | February,1999 |

# Introduction

The *nested logit* (NL) model is the preferred specification of a discrete choice model when analysts move beyond the multinomial logit (MNL) model. [See Ortuzar (1998) for an historical perspective on nested logit models.] Despite the increasing availability of other less restrictive models (in terms of the nature of the random components of the utility expressions for each alternative) such as *heteroscedastic extreme value*, *mixed logit*, *random parameter logit*, *covariance heterogeneity logit* and *multinomial probit* - see Louviere et al (1999 Chapter 6 Appendix B) for a review - there remain reasons why the nested logit (NL) model will continue to be estimated. For example, the NL model is relatively easy to estimate, and, with its closed-form structure, it is easy to implement in the simulation of market shares before and after a policy change.

Specialists involved in the development of NL models, especially the active set of individuals researching estimation methods and developing software, have recently entered into a dialogue on the model specifications required in using software to ensure that the estimation is consistent with utility maximisation, and how one should handle degenerate branches (i.e those with only a single alternative). Much of the discussion has taken place by email, however the sentiment of the dialogue is partially represented in a series of recent papers by Koppelman and Wen (1998a, 1998b) and Hunt (1998). The objective of this note is to gather the presentation into a single transparent notation and to illustrate how one sets up an NL model to obtain outputs consistent with McFadden's NL model for utility maximization, a derivative of his Generalised Extreme Value (GEV) model [McFadden (1981)].

# A Common Notation for Nested Logit Models

We propose the following notation as a method of unifying the different forms of the NL model. Each observed (or representative) component of the utility expression for an alternative (usually denoted as $V_k$ for the $k$th alternative) is defined in terms of four items – the parameters associated with each explanatory variable, an alternative-specific constant, a scale parameter and the explanatory variables. The utility of alternative $k$ for individual $t$ is

$$
\begin{aligned}
U_{tk} &= g_k(\alpha_k,\, \mathbf{b}'\mathbf{x}_{tk},\, \varepsilon_{tk}) \\
&= g_k(V_{tk}, \varepsilon_{tk}) \\
&= \alpha_k + \mathbf{b}'\mathbf{x}_{tk} + \varepsilon_{tk}, \\
\mathrm{Var}[\varepsilon_{tk}] &= \sigma^2 = \kappa/\theta^2.
\end{aligned}
$$

The scale parameter, $\theta$, is proportional to the inverse of the standard deviation of the random component in the utility expression, $\sigma$, and is a critical input into the set up of the NL model [Ben-Akiva and Lerman (1985), Louviere et al. (1999)]. Under the assumptions now well established in the literature, utility maximization in the presence of random components which have independent (across choices and individuals) extreme value distributions produces a simple closed form for the probability that choice $k$ is made;

$$\text{Prob}[U_{tk} > U_{tj} \; \forall \; j \neq k] = \frac{\exp(\alpha_k + \mathbf{b'x}_{tk})}{\sum_{j=1}^{K} \exp(\alpha_j + \mathbf{b'x}_{tj})} \; .$$

Under these assumptions, the common variance of the assumed i.i.d. random components is lost. The same observed set of choices emerges regardless of the (common) scaling of the utilities. Hence the latent variance is normalized at one, not as a restriction, but of necessity for identification.

One justification for moving from the MNL model to an NL model is to recognize (or at least test for) the possibility that the standard deviations (or variances) of the random error components in the utility expressions are different across groups of alternatives in the choice set. This arises because the sources of utility associated with the alternatives are not fully accommodated in $V_k$. The missing sources of utility may differentially impact on the random components across the alternatives, resulting in different variances. To accommodate the possibility of differential variances, we must explicitly introduce the scale parameters into each of the utility expressions. (If all scale parameters are equal, then the NL model 'collapses' back to a simple MNL model.) Hunt (1998) discusses the underlying conditions that produce the nested logit model as a result of utility maximization within a partitioned choice set.

The notation for a three-level nested logit model covers the majority of applications. The literature suggests that very few analysts estimate models with more than three levels, and two levels are the most common. However it will be shown below that a two-level model *may* require a third level (in which the lowest level is a set of dummy nodes and links) simply to ensure consistency with utility maximization (which has nothing to do with a desire to test a three level NL model). It is also common for a nested structure to have a branch with only one alternative. This is referred to as a degenerate branch. This requires careful definition in estimation. We will return to this point below.

It is useful to represent each level in an NL tree by a unique descriptor. For a three level tree (Figure 1), the top level will be represented by *limbs*, the middle level by a number of *branches* and the bottom level by a set of *elemental alternatives*, or *twigs*. We have $k=1,\ldots,K$ elemental alternatives, $j=1,\ldots,J$ branch composite alternatives and $i=1,\ldots,I$ limb composite alternatives. We use the notation $k|j,i$ to denote alternative $k$ in branch $j$ of limb $i$ and $j|i$ to denote branch $j$ in limb $i$.
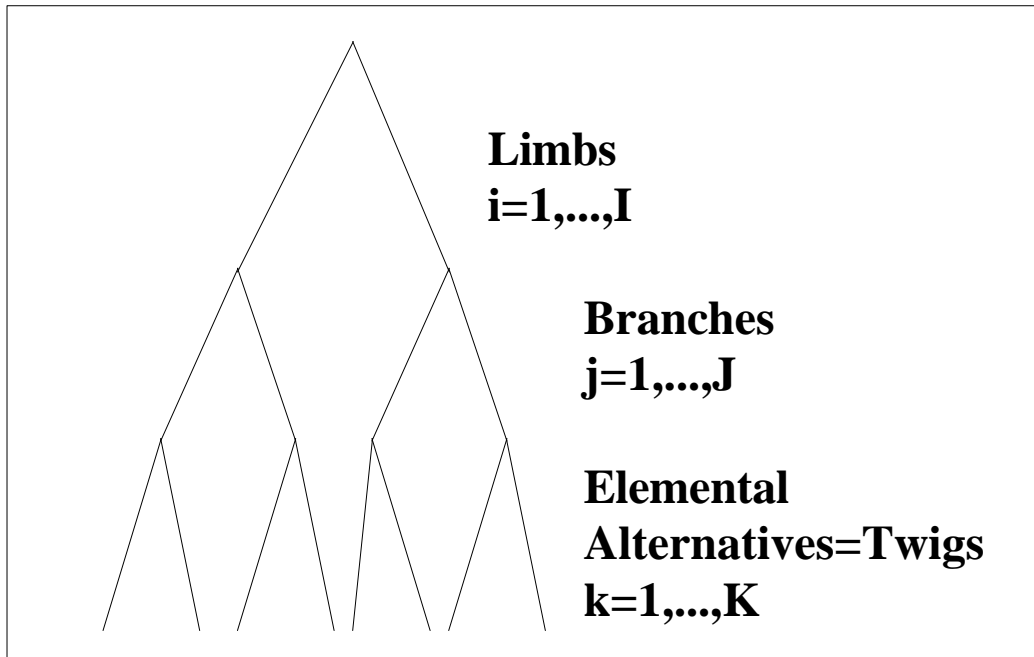
**Figure 1    Descriptors for a three-level NL tree**

Define parameter vectors in the utility functions at each level as follows: **b** for elemental alternatives, **a** for branch composite alternatives, and **c** for limb composite alternatives. The branch level composite alternative involves an aggregation of the lower level alternatives.  As discussed below, a branch specific scale parameter $\mu(j|i)$ will be associated with the lowest level of the tree.  Each elemental alternative in the $j$'th branch will actually have scale parameter $\mu'(k|j,i)$.  Since these will, of necessity, be equal for all alternatives in a branch, the distinction by $k$ is meaningless.  As such, we collapse these into $\mu(j|i)$.  The parameters $\lambda(j|i)$ will be associated with the branch level. The inclusive value (IV) parameters at the branch level will involve the ratios $\lambda(j|i)/\mu(j|i)$.  For identification, it will be necessary to normalise one of these parameters, either $\mu(j|i)$ or $\lambda(j|i)$ to one.   The inclusive value (IV) parameters associated with the composite alternative at each level are thus defined either by scale parameters $\mu(j|i)$ or $\lambda(j|i)$ for branches, and $\gamma(i)$ for limbs. The IV parameters associated with the IV variable in a branch, calculated from the natural logarithm of the sum of the exponential of the $V_k$ expressions at the elemental alternative level directly below a branch, implicitly have associated parameters defined as the $\lambda(j|i)/\mu(j|i)$, but, as noted, some normalisation is required. Some analysts do this without acknowledgment of which normalisation they have used, which makes the comparison of reported results between studies difficult. Normalisation is simply the process of setting one or more scale parameters equal to unity, while allowing the other scale parameters to be estimated.

The literature is vague on the implications of choosing the normalisation of $\lambda(j|i) = 1$ versus $\mu(j|i) = 1$. It is important to note that the notation $\mu'(m|j,i)$ used below refers to the scale parameter for each elemental alternative. However, since a nested logit

structure is specified to test for the presence of identical scale within a subset of alternatives, it comes as no surprise that all alternatives partitioned under a common branch have the same scale parameter imposed on them. Thus $\mu'(k|j,i) = \mu(j|i)$ for every $k=1,\ldots,K|j,i$ alternatives in branch $j$ in limb $i$.

We now set out the probability choice system (PCS) for the situation where we normalise on $\mu(j|i)$ – called Random Utility Model 1 (RU1), and the situation where we normalise on $\lambda(j|i)$ – called Random Utility Model 2 (RU2). We ignore the subscripts for an individual. For later purposes, we now define the three-level PCS,

$$P(k,j,i) = P(k|j,i) \times P(j|\text{i}) \times P(i).$$

## *Random utility model 1 (RU1)*

The choice probabilities for the *elemental alternatives* are defined as:

$$P(k \mid j,i) = \frac{\exp[\mathbf{b}'\mathbf{x}(k \mid j,i)]}{\sum_{l=1}^{K|j,i} \exp[\mathbf{b}'\mathbf{x}(l \mid j,i)]} = \frac{\exp[\mathbf{b}'\mathbf{x}(k \mid j,i)]}{\exp[IV(j \mid i)]}$$

where $k|j,i$ = elemental alternative $k$ in branch $j$ of limb $i$, $K|j,i$ = number of elemental alternatives in branch $j$ of limb $i$, and the inclusive value for branch $j$ in limb $i$ is

$$IV(j \mid i) = \log \sum_{k=1}^{K|j,i} \exp[\mathbf{b}'\mathbf{x}(k \mid j,i)]$$

The *branch level* probability is

$$p(j \mid i) = \frac{\exp\{\boldsymbol{l}(j \mid i)[\mathbf{a}'\mathbf{y}(j \mid i) + IV(j \mid i)]\}}{\sum_{m=1}^{J|i} \exp\{\boldsymbol{l}(m \mid i)[\mathbf{a}'\mathbf{y}(m \mid i) + IV(m \mid i)]\}} = \frac{\exp\{\boldsymbol{l}(j \mid i)[\mathbf{a}'\mathbf{y}(j \mid i) + IV(j \mid i)]\}}{\exp[IV(i)]}$$

where $j|i$ = branch $j$ in limb $i$, $J|i$ = number of branches in limb $i$, and

$$IV(i) = \log \sum_{j=1}^{J|i} \exp\{\boldsymbol{l}(j \mid i)[\mathbf{a}'\mathbf{y}(j \mid i) + IV(j \mid i)]\}$$

Finally, the *limb level* is defined by

$$p(i) = \frac{\exp\{\boldsymbol{g}(i)[\mathbf{c}'\mathbf{z}(i) + IV(i)]\}}{\sum_{n=1}^{I} \exp\{\boldsymbol{g}(n)[\mathbf{c}'\mathbf{z}(n) + IV(n)]\}} = \frac{\exp\{\boldsymbol{g}(i)[\mathbf{c}'\mathbf{z}(i) + IV(i)]\}}{\exp(IV)}$$

where $I$ = number of limbs in the three level tree and

$$IV = \log \sum_{i=1}^{I} \exp\{\mathbf{g}(i)[\mathbf{c}'\,\mathbf{z}(i) + IV(i)]\}$$

RU1 has been described (e.g. by Bates 1999) as corresponding to a non-normalised nested logit (NNNL) specification, since the parameters are scaled at the lowest level (i.e. for $\mu'(k|j,i) = \mu(j|i) = 1$.

## Random utility model 2 (RU2)

Suppose, instead, we normalise the upper level parameters and allow the lower level scale parameters to be free. The *elemental alternatives level* probabilities will be:

$$P(k|j,i) = \frac{\exp\left[\mu'(k|j,i)\mathbf{b}'\,\mathbf{x}\,(k|j,i)\right]}{\displaystyle\sum_{l=1}^{K|j,i} \exp\left[\mu'(l|j,i)\mathbf{b}'\,\mathbf{x}\,(l|j,i)\right]} = \frac{\exp\left[\mu'(k|j,i)\mathbf{b}'\,\mathbf{x}\,(k|j,i)\right]}{\exp[IV(j|i)]}$$

$$= \frac{\exp[\mu(j|i)\mathbf{b}'\,\mathbf{x}\,(k|j,i)]}{\displaystyle\sum_{l=1}^{K|j,i}\exp[\mu(j|i)\mathbf{b}'\,\mathbf{x}\,(l|j,i)]} = \frac{\exp[\mu(j|i)\mathbf{b}'\,\mathbf{x}\,(k|j,i)]}{\exp[IV(j|i)]}\,.$$

[with the latter equality resulting from the identification restriction $\mu'(k|j,i) = \mu'(m|j,i) = \mu(j|i)$] and

$$IV(j|i) = \log \sum_{k=1}^{K|j,i}\exp[\mathbf{m}(j|i)\mathbf{b}'\,\mathbf{x}(k|j,i)]$$

The *branch level* is defined by:

$$p(j \mid i) = \cfrac{\exp\left\{\boldsymbol{g}(i)\left[\mathbf{a'}\,\mathbf{y}(j \mid i) + \cfrac{1}{\boldsymbol{m}(j \mid i)}\,IV(j \mid i)\right]\right\}}{\sum\limits_{m=1}^{J\mid i}\exp\left\{\boldsymbol{g}(i)\left[\mathbf{a'}\,\mathbf{y}(m \mid i) + \cfrac{1}{\boldsymbol{m}(m \mid i)}\,IV(m \mid i)\right]\right\}} = \cfrac{\exp\left\{\boldsymbol{g}(i)\left[\mathbf{a'}\,\mathbf{y}(j \mid i) + \cfrac{1}{\boldsymbol{m}(j \mid i)}\,IV(j \mid i)\right]\right\}}{\exp[IV(i)]}$$

$$IV(i) = \log\sum_{j=1}^{J\mid i}\exp\left\{\boldsymbol{g}(i)\left[a'\,y(j \mid i) + \cfrac{1}{\boldsymbol{m}(j \mid i)}\,IV(j \mid i)\right]\right\}$$

The *limb level* is defined by :

$$p(i) = \cfrac{\exp\left[\mathbf{c'}\,\mathbf{z}(i) + \cfrac{1}{\boldsymbol{g}(i)}\,IV(i)\right]}{\sum\limits_{n=1}^{I}\exp\left[\mathbf{c'}\,\mathbf{z}(n) + \cfrac{1}{\boldsymbol{g}(n)}\,IV(n)\right]} = \cfrac{\exp\left[\mathbf{c'}\,\mathbf{z}(i) + \cfrac{1}{\boldsymbol{g}(i)}\,IV(i)\right]}{\exp(IV)}$$

$$IV = \log\sum_{i=1}^{I}\exp\left[c'\,z(i) + \cfrac{1}{\boldsymbol{g}(I)}\,IV(i)\right]$$

It is typically assumed that it is arbitrary as to which scale parameter is normalised. [See Hunt (1998) for a useful discussion.] Most applications normalize the scale parameters associated with the branch level utility expressions [ie $\lambda(j|i)$] at 1 as in RU2 above, then allow the scale parameters associated with the elemental alternatives [$\mu(j|i)$] and hence the inclusive value parameters in the branch composite alternatives to be unrestricted. It is implicitly assumed that the empirical results are identical to those that would be obtained if RU1 were instead the specification (even though parameter estimates are numerically different). *But, within the context of a two-level partition of a nest estimated as a two-level model, unless all attribute parameters are alternative-specific, this assumption is only true if the non-normalised scale parameters are constrained to be the same across nodes within the same level of a tree (i.e., at the branch level for two levels, and at the branch level and the limb level for three-levels).* This latter result actually appears explicitly in some early studies of this model, e.g., Maddala (1983, p.70) and Quigley (1985), but is frequently ignored in more recent applications. Note that in the common case of estimation of RU2 with two levels (which eliminates $\gamma(i)$) the 'free' IV parameter estimated will typically be $1/\mu(j|i)$. Other interpretations of this result are discussed in Hunt (1998).

## Conditions to ensure consistency with utility maximization

The previous section set out a uniform notation for a three-level NL model, choosing a different level in the tree for normalisation (ie. setting scale parameters to an arbitrary value, typically unity). We have chosen levels one and two respectively for the RU1 and RU2 models. We now are ready to present a range of alternative empirical specifications for the NL model, some of which satisfy utility maximization either

directly from estimation or by some simple transformation of the estimated parameters. Compliance with utility maximization requires that the addition of a constant value to all elemental alternatives has no effect on the choice probabilities of the alternatives [McFadden (1981)]. We limit the discussion to a two-level NL model and initially assume that all branches have at least two elemental alternatives. The important case of a degenerate branch (i.e., only one elemental alternative) is treated separately later.

## An Empirical Illustration

To investigate the implications of alternative model specifications, we have estimated nine two-level models, using data collected in 1986 on non-business interurban trips between Sydney, Canberra and Melbourne. A total of 210 travellers chose a mode of transport from four alternatives – plane, car, train and bus. Details of the data are provided in Econometric Software (1998) and Louviere et al (1999). The utility functions for the four alternatives are specified as follows:

$$U_{Train} = \alpha_T + \beta GC + \gamma TC + \delta Ttime + \varepsilon_T$$
$$U_{Bus} = \alpha_B + \beta GC + \gamma TC + \delta Ttime + \varepsilon_B$$
$$U_{Plane} = \alpha_P + \beta GC + \gamma TC + \delta Ttime + \varepsilon_P$$
$$U_{Car} = \beta GC + \varepsilon_C$$

The variables in the utility functions in addition to the alternative specific constants are

GC = Generalized cost (in dollars)
= out-of-pocket fuel cost for car or fare for plane, train and bus + time cost

TC = time cost
= product of linehaul travel time and value of travel time savings in $/minute.

Ttime = Transfer time (in minutes)
= the time spent waiting for and transferring to plane, train, bus.

Table 1 presents full information maximum likelihood (FIML) estimates of a two-level non-degenerate NL model. The tree structure for Table 1 has two branches, PUBLIC = (Train,Bus) and OTHER = (Car,Plane). In the PCS for this model, household income enters the probability of the branch choice directly in the utility for OTHER. Inclusive values from the lowest level enter both utility functions at the branch level. Table 2 presents FIML estimates of a two-level partially degenerate NL model. The tree structure for the models in Table 2, save for Model 7 which has an artificial third level, is FLY(Plane) and GROUND(Train,Bus,Car).

Estimates for both the non-normalised nested logit (NNNL) model and the utility maximising (GEV-NL) parameterizations are presented. In the case of the GEV model parameterisation, estimates under each of the two normalizations (RU1: $\mu=1$ and RU2: $\lambda=1$) are provided as are estimates with the IV parameters restricted to equality within a level of the tree and unrestricted.

Eight models are summarised in Table 1 and six models in Table 2. Since there is only one limb, we drop the limb indicator from $\lambda(j|i)$ and denote it simply as $\lambda(j)$.

Model 1:  RU1 with scale parameters equal within a level $[\lambda(1)=\lambda(2)]$;

Model 2:  RU1 with scale parameters unrestricted within a level $[\lambda(1) \neq \lambda(2)]$;

Model 3:  RU2 with scale parameters equal within a level (not applicable for a degenerate branch) $[\mu(1) = \mu(2)]$

Model 4:  RU2 with scale parameters unrestricted within a level $[\mu(1) \neq \mu(2)]$

Model 5:  Non-normalised NL model with dummy nodes and links to allow unrestricted scale parameters in the presence of generic attributes to recover parameter estimates that are consistent with utility maximisation. This is equivalent up to scale with RU2 (model 4).

Model 6:  Non-normalised NL model with no dummy nodes/links and different scale parameters within a level. This is a typical NL model implemented by many practictioners (and is equivalent to RU1 (Model 2)).

Model 7: RU2 with unrestricted scale parameters and dummy nodes and links to comply with utility maximisation (for partial degeneracy). Since Model 7 is identical to Model 8 in Table 1, it is not presented. - Table 2 only:

Models 8 and 9: For the non-degenerate NL model (Table 1), these are RU1 and RU2 in which all parameters are alternative-specific and scale parameters are unrestricted across branches.

All results reported in Tables 1 and 2 are obtained using LIMDEP Version 7 (Revised, December 1998) [Econometric Software (1998)]. The IV parameters for RU1 and RU2 that LIMDEP reports are the $\mu$s and the $\lambda$s that are shown in the equations above. These $\mu$s and $\lambda$s are proportional to the reciprocal of the standard deviation of the random component. The *t*-values in parenthesis for the NNNL model require correction to compare with RU1 and RU2. Koppelman and Wen (1998b) provide the procedure to adjust the t-values. For a two-level model, the corrected variance and hence standard error of estimate for the NNNL model is:

$$Var\left[\hat{\beta}_{RU}\right] = \beta_{NN}^2 Var\left[\hat{\mu}_{NN}\right] + \mu_{NN}^2 Var\left[\hat{\beta}_{NN}\right] + 2\mu_{NN}\beta_{NN} Cov\left[\hat{\mu}_{NN}, \hat{\beta}_{NN}\right]$$

*The case of generic attribute parameters*

Beginning with the non-degenerate case, it can be seen in Table 1 that the GEV parameterization estimates with IV parameters unrestricted (Models 2 and 4) are *not* invariant to the normalisation chosen. Not only is there no obvious relationship between the two sets of parameter estimates, the log-likelihood function values at convergence are not equal (–184.31 vs. –188.43). When the GEV parameterisation is estimated subject to the restriction that the IV parameters be equal (Models 1 and 3), invariance is achieved across normalisation after accounting for the difference in scaling. The log-likelihood function values at convergence are equal (-190.178), and the IV parameter estimates are inverses of one another (1/0.773 = 1.293, within rounding error). Multiplying the utility function parameter estimates at the elemental alternatives level (i.e. $\alpha_{Plane}$, $\alpha_{Train}$, $\alpha_{Bus}$, GC, TTME) by the corresponding IV parameter estimate in one normalisation (eg Model 1) yields the utility function parameter estimates in another normalisation (eg Model 3). For example, in Model 3, (1/1.293)×5.873 for Train constant = 4.542 in Model 1.

**Table 1. Summary of Alternative Model Specifications for a Non-degenerate NL Model Tree structure: other {plane, car} vs public transport {train, bus} except for model 5 which is OTHER {planem (plane), carm(car)} vs PUBLIC TRANSPORT {trainm (train), busm (bus)}**

*Note: there is no model 7 in order to keep equivalent model numbering in Tables 1 and 2.*

| Variables | Alternative | Model 1: RU1 | Model 2: RU1 | Model 3: RU2 | Model 4: RU2 | Model 5: NNNL** | Model 6: NNNL** | Model 8: RU1* | Model 9: RU2 |
|---|---|---|---|---|---|---|---|---|---|
| Train constant | Train | 4.542 (6.64) | 3.757 (5.8) | 5.873 (5.8) | 6.159 (5.7) | 3.6842 (2.34) | 3.757 (5.8) | 17.396 (2.2) | 2.577 (3.8) |
| Bus constant | Bus | 3.924 (5.83) | 2.977 (4.4) | 5.075 (6.0) | 5.380 (5.8) | 3.218 (2.18) | 2.977 (4.4) | 19.523 (2.4) | 2.892 (3.73) |
| Plane constant | Plane | 5.0307 (6.81) | 4.980 (6.7) | 6.507 (5.7) | 6.154 (5.2) | 3.681 (2.34) | 4.980 (6.7) | 4.165 (3.3) | 4.165 (3.5) |
| Generalised Cost ($) | All | -.01088 (-2.6) | -.0148 (-3.5) | -.01407 (-2.6) | -.01955 (-3.2) | -.01169 (-1.8) | -.0148 (-3.5) | | |
| Transfer time (mins.) | All excl car | -.0859 (-7.4) | -.0861 (-7.3) | -.1111 (-5.5) | -.1064 (-5.2) | -.0637 (-2.5) | -.0861 (-7.3) | | |
| Hhld income ($000s) | other | .03456 (3.2) | .0172 (2.9) | .0447 (4.0) | .0426 (3.8) | .0426 (3.8) | .0416 (3.4) | .04269 (4.24) | .04269 (3.8) |
| Generalised Cost ($) | Air | | | | | | | .00492 (.56) | .00492 (.60) |
| Generalised Cost ($) | Train | | | | | | | -.0943 (-3.0) | -.0139 (-2.8) |
| Generalised Cost ($) | Bus | | | | | | | -.1065 (-2.9) | -.0158 (-2.9) |
| Generalised Cost ($) | Car | | | | | | | -.0143 (-2.5) | -.0143 (-3.0) |
| Transfer Time (mins) | Air | | | | | | | -.1048 (-6.2) | -.1048 (-7.1) |
| Transfer Time (mins) | Train | | | | | | | -.0787 (-2.5) | -.0116 (-1.9) |
| Transfer Time (mins) | Bus | | | | | | | -.1531 (-3.5) | -.0227 (-1.9) |
| *Inclusive Value* | | | | | | | | | |
| *Parameters:* | | | | | | | | | |
| IV Other | Other | 1.293 (5.3) | 2.42 (4.6) | .773 (3.8) | 0.579 (3.3) | 0.969 (3.2)** | 2.42 (4.6) | 1.00 (fixed) | 1.00 (fixed) |
| IV Public transport | Public Transport | 1.293 (5.3) | 1.28 (5.1) | .773 (3.8) | 1.03 (3.2) | 1.724 (3.3)** | 1.28 (5.1) | .148 (2.0) | 6.75 (1.9) |
| Log-likelihood | | -190.178 | -184.31 | -190.178 | -188.43 | -188.43 | -184.31 | -177.82 | -177.82 |
| *Direct Elasticities:* | | | | | | | | | |
| | Plane | -.544 | -.797 | -.544 | -.666 | -.666 | -.797 | .228 | .228 |
| | Car | -.651 | -1.081 | -.651 | -.762 | -.762 | -1.081 | -.799 | -.799 |
| | Train | -.629 | -.854 | -.629 | -.910 | -.910 | -.854 | -.188 | -1.27 |
| | Bus | -.759 | -1.014 | -.759 | -1.174 | -1.174 | -1.014 | -.343 | -2.32 |

Model 8 with all alternative-specific attributes produces the exact parameter estimates, overall goodness of fit and elasticities as the NNNL model (and hence it is not reported).

** standard errors are uncorrected.

*** = IV parameters in Model 5 based on imposing equality of IV for (other, trainm, busm) and for (public transport, planem,carm).

**Table 2. Summary of Alternative Model Specifications for a Partial Degeneracy NL Model Tree structure: fly {plane} vs ground {train, bus, car}**

Note: Model 3 is not defined for a degenerate branch model when the IV parameters are forced to equality. Forcing a constraint on Model 4 (ie equal IV parameters) to obtain Model 3 produced exactly the same results for all the parameters. This is exactly what should happen. Since the IV parameter is not identified, no linear constraint that is imposed that involves this parameter is binding. Model 7 tree is *other{fly (plane) vs auto (car)} vs Land PT{ public transport (train, bus}).*

| *Variables* | *Alternatives* | *Model 1: RU1* | *Model 2: RU1* | *Model 4: RU2* | *Model5:NNNL\** | *Model6:NNNL\** | *Model 7: RU2* |
|---|---|---|---|---|---|---|---|
| Train constant | Train | 5.070 (7.6) | 5.065 (7.7) | 2.622 (5.9) | 9.805 (3.4) | 5.065 (7.7) | 4.584 (2.7) |
| Bus constant | Bus | 4.145 (6.7) | 4.096 (6.7) | 2.143 (5.5) | 8.015 (3.3) | 4.096 (6.7) | 3.849 (2.7) |
| Plane constant | Plane | 5.167 (4.3) | 6.042 (5.04) | 2.672 (3.0) | 9.993 (3.8) | 6.042 (5.04) | 5.196 (2.7) |
| Generalised Cost ($) | All | -.0291 (-3.6) | -.0316 (-3.9) | -.0151 (4.32) | -.0564 (2.8) | -.0316 (-3.9) | -.0187 (-2.4) |
| Transfer time (min.) | All excl car | -.1156 (8.15) | -.1127 (-7.9) | -.0598 (-5.9) | -.2236 (3.7) | -.1127 (-7.9) | -.0867 (-2.7) |
| Hhld income ($000s) | fly | .02837 (1.4) | .0262 (1.5) | .0143 (1.35) | .01467 (1.4) | .01533 (1.6) | .02108 (1.4) |
| Party Size | auto | | | | | | .4330 (2.0) |
| *Inclusive Value* | | | | | | | |
| *Parameters:* | | | | | | | |
| IV | Fly | .517 (4.1) | .586 (4.2) | 1.00(.67E+15)\*\* | 1.00(.67E+15)\*\* | .586 (4.2) | Not applicable |
| IV | Ground | .517 (4.1) | .389 (3.1) | 1.934 (5.0)\*\* | .517 (5.0) | .389 (3.1) | |
| IV | Auto | | | | | | Not applicable |
| IV | Public Transport | | | | | | 1.26 (2.3) |
| IV | Other | | | | | | 0.844 (2.1) |
| IV | Land PT | | | | | | Not applicable |
| Log-likelihood | | -194.94 | -193.66 | -194.94 | -194.94 | -193.66 | -194.27 |
| *Direct Elasticities:* | | | | | | | |
| | Plane | -.864 | -1.033 | -.864 | -.864 | -1.033 | -.859 |
| | Car | -1.332 | -1.353 | -1.332 | -1.332 | -1.353 | -.946 |
| | Train | -1.317 | -1.419 | -1.317 | -1.317 | -1.419 | -1.076 |
| | Bus | -1.650 | -1.878 | -1.650 | -1.650 | -1.878 | -1.378 |

\* standard errors are uncorrected.

\*\* = IV parameters in Model 5 based on imposing equality of IV for (ground, airm). It is not applicable for (fly, trainm, busm, carm) since fly is degenerate.

The points made above about invariance, or the lack of it, scaling, and the equivalence of GEV and NNNL under the appropriate set of parametric restrictions are also illustrated in Table 2 for the case of a partially degenerate NL model structure. However, an additional and important result emerges for the partial degeneracy case. If the IV parameters are unrestricted, the GEV model "estimate" of the parameter on the degenerate partition IV is unity under the $\lambda=1$ normalisation. This will always be the case because of the cancellation of the IV parameter and the lower-level scaling parameter in the GEV model in the degenerate partition. The results will be invariant to whatever value this parameter is set to. To see this, consider the results for the unrestricted GEV model presented as Model 4 in Table 2. The IV parameter is "estimated" to be 1.934, and if we were to report Model 3, all of the other estimates would be the same as in Model 4 and the log-likelihood function values at convergence are identical (-194.94). In a degenerate branch, whatever the value of $(1/\mu)$, it will cancel with the lower-level scaling parameter, $\mu$, in the degenerate partition marginal probability. If we select $\mu=1$ for normalisation (in contrast to $\lambda$) in the presence of a degenerate branch, the results will produce restricted (Model 1) or unrestricted (Model 2) estimates of $\lambda$ which, unlike $\mu$, do not cancel out in the degenerate branch. [Hunt (1998) pursues this issue at length.]

To illustrate the equivalence of behavioral outputs for RU1 and RU2, Tables 1 and 2 present the weighted aggregate direct elasticities for the relationship between the generalised cost of alternative *kji* and the probability of choosing alternative *k|ji*. As expected the results are identical for RU1 (Model 1) and RU2 (Model 3) when the IV parameters are equal across all branches at a level in the GEV model. The elasticities are significantly different from those obtained from Models 2 and 4, although Models 4 and 5 produce the same results (see below). Model 6 (equivalent to Model 2) is a common model specification in which parameters of attributes are generic and scale parameters are unrestricted within a level of the NL model with no constraints imposed to recover the utility-maximisation estimates.

## *Allowing different scale parameters across nodes in a level in the presence of generic and/or alternative-specific attribute parameters between partitions*

When we allow the IV parameters to be unrestricted in the RU1 and RU2 GEV models and in the NNNL model we fail to comply with normalisation invariance, and for models 2 and 6 we also fail to produce consistency with utility maximisation. RU1 (Model 2) fails to comply with utility maximisation because of the absence of explicit scaling in the utility expressions for elemental alternatives. We obtain different results on overall goodness of fit and the range of behavioural outputs such as elasticities.

For a given nested structure and set of attributes there can only be one utility maximising solution. This presents a dilemma, since we often want the scale parameters to vary between branches and/or limbs or at least test for non-equivalence. This is after all, the main reason why we seek out alternative nested structures. Fortunately there is a solution, depending on whether one opts for a specification in

which some or all of the parameters are generic or whether they are all alternative-specific. Models 5 to 9 are alternative specifications.

If all attributes between partitions are unrestricted (ie alternative-specific), unrestricted scale parameters are compliant with utility maximisation under all specifications (ie RU2=1, RU2 and NNNL). Intuitively, the fully alternative-specific specification avoids any artificial 'transfer' of information from the attribute parameters to the scale parameters that occurs when restrictions are imposed on parameter estimates. Models 8 and 9 in Table 1 are totally alternative-specific. The scale parameters for models 8 and 9 are the inverse of each other. That is, for the unrestricted IV, 0.148 in model 8 equals (1/6.75) in model 9. The alternative-specific parameter estimates associated with attributes in the public transport branch for model 8 can be recovered from model 9 by a scale transformation. For example, .148* 17.396 for the train constant equals 2.577 in model 9. The estimated parameters are identical in models 8 and 9 for the 'other' modes since their IV parameter is restricted to equal unity in both models. This demonstrates the equivalence up to scale of RU1 and RU2 when all attribute parameters (including IV) are unrestricted.

When we impose the generic condition on an attribute associated with alternatives in different partitions of the nest, Koppelman and Wen (1998a,b) (and Daly in advice to ALOGIT subscribers) have shown how one can recover compliance with utility maximisation in an NNNL model under the unrestricted scale condition (within a level of the NL model) by adding dummy nodes and links below the bottom level and imposing cross-branch equality constraints as illustrated in Figure 2. Intuitively, what we are doing is allowing for differences in scale parameters at each branch but preserving the (constant) ratio of the IV parameters between two levels through the introduction of the scale parameters at the elemental level; the latter requiring the additional lower level in an NNNL specification. The NNNL specification does not allow unrestricted values of scale at the elemental level, unlike RU2, for example. Preserving a constant ratio through cross-over equality constraints between levels in the nest satisfies the necessary condition of choice probability invariance to the addition of a constant in the utility expression of all elemental alternatives.
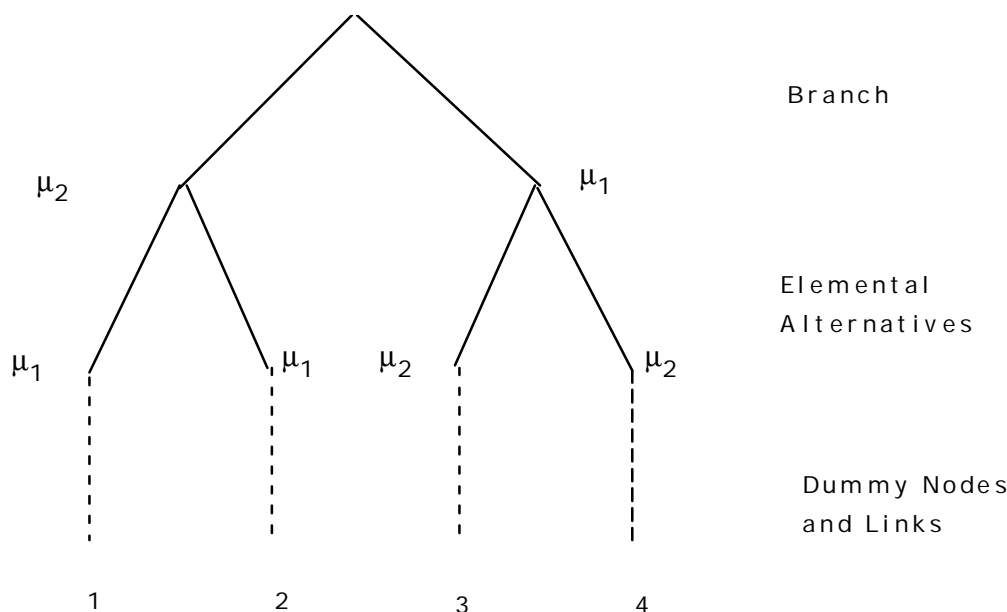
**Figure 2**     **Estimating a Two-level Model to allow for Unrestricted Scale Parameters within a Level**

Adding an extra level is not designed to investigate the behavioural implications of a three-level model; rather it is a 'procedure' to reveal the scale parameters at upper levels where they have not been identified. This procedure is fairly straightforward for two branches (see Model 5 in Tables 1 and 2). With more than two branches, one has to specify additional levels for each branch. The number of levels grows quite dramatically. However, there is one way of simplifying this procedure, if we recognise that the ratio of the scale parameters between adjacent levels must be constant. Thus, for any number of branches, consistency with utility maximisation requires that the product of all the ratios of scale parameters between levels must be identical from the root to all elemental alternatives. To facilitate this, one can add a single link below each real alternative with the scale of that link set equal to the product of the scales of all scale parameters not included in the path to that alternative. For example, in the case of three branches with scales equal to $\lambda_1$, $\lambda_2$ and $\lambda_3$; the scale below the first branch would be ($\lambda_2*\lambda_3$), below the second branch it would be ($\lambda_1*\lambda_3$) and below the third branch it would be ($\lambda_1*\lambda_2$).

Model 5 is estimated as an NNNL model with the addition of a lower level of nodes and links with cross-branch equality constraints on the scale parameters. For example, in Table 1, the tree structure is as follows: {Other [planem (plane), carm (car)], Public Transport [trainm (train), busm (bus)]}. The cross-over constraint for two branches sets the scale parameters to equality for {*Other*, trainm, busm} and {*Public Transport*, planem, carm}. The Model 5 (Table 1) produces results which are identical to RU2 (model 4) in respect of goodness-of-fit and elasticities; with all parameter estimates equivalent up to scale. Since we have two scale parameters in Model 5, the ratio of each branches IV parameters to their equivalent in Model 4 provides the adjustment factor to translate Model 5 parameters into Model 4 parameters (or vice versa). For

14

example, the ratio of .579/.969 = 1.03/1.724 = 0.597. If we multiply the train specific constant in Model 4 of 6.159 by 0.597, we obtain 3.6842, the train-specific constant in Model 5. This is an important finding, because it indicates that:

> *the application of the RU2 specification with unrestricted scale parameters in the presence of generic parameters across branches for the attributes is identical to the results obtained by estimating the NNNL model with an extra level of nodes and links.*

RU2 thus avoids the need to introduce the extra level.[1] The equivalent findings are shown in Table 2 where the scale ratio is 3.74. Intuitively one might expect such a result given that RU2 allows the scale parameters to be freely estimated at the lower level (in contrast to RU1 where they are normalised to 1.0). One can implement this procedure under an exact RU2 model specification to facilitate situations where one wishes to allow scale parameters at a level in the nest to be different across branches in the presence or absence of a generic specification of attribute parameters. *The estimation results in Model 4 are exactly correct and require no further adjustments.* The procedure can also be implemented under an NNNL specification (with an extra level of nodes and links) (Model 5). The elasticities, marginal rates of substitution, goodness-of fit are identical in Models 4 and 5. The parameter estimates are identical up to the ratio of scales.

## Conclusions

The empirical applications and discussion has identified the model specification required to ensure compliance with the necessary conditions for utility maximisation.

This can be achieved for a *GEV-NL model* by either

- setting the IV parameters to be the same at a level in the nest in the presence of generic parameters, or
- implementing  the RU2 specification and allowing the IV parameters to be free in the presence of generic attribute parameters between partitions of a nest, or
- setting all attribute parameters to be alternative-specific between partitions, allowing IV parameters to be unrestricted.

This can be achieved for a *Non-normalised NL model* by either

- setting the scale parameters to be the same at a level in the nest (for the non-normalised scale parameters) and rescaling all estimated parameters associated with elemental alternatives by the estimated IV parameter, or
- allowing the IV parameters to be free, and adding an additional level at the bottom of the tree through dummy nodes and links, and constraining the scale parameters

---

[1] From a practical perspective, this enables programs like Limdep that limit the number of levels which can be jointly estimated to use all levels for real behavioural analysis.

at the elemental alternatives level to equal those of the dummy nodes of all other branches in the total NL model, or

- setting all attribute parameters to be alternative-specific between partitions, allowing IV parameters to be unrestricted.

# References

Bates, J. (1999) "More Thoughts on Nested Logit," John Bates Services, Oxford (mimeo), January

Ben-Akiva, M. and Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge.

Daly, A. (1987) Estimating 'tree' logit models, *Transportation Research*, 21B(4), 251-267.

Econometric Software (1998) *Limdep Version 7 for Windows*, Econometric Software, New York and Sydney, December revision.

Hague Consulting Group (1995) *ALOGIT: User's Guide*, Hague Consulting Group, Den Haag.

Hunt, G.L. (1998) Nested logit models with partial degeneracy, *Department of Economics, University of Maine*, December (revised).

Koppelman, F.S. and Wen, C.H. (1998a) Alternative nested logit models: structure, properties and estimation, *Transportation Research* 32B(5), June, 289-298.

Koppelman, F.S. and Wen, C.H. (1998b) Nested logit models: which are you using? *Transportation Research Record* ,1645, 1-7.

Louviere, J.J., Hensher, D.A. and Swait, J. (1999) *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*, 614pp

Maddala, J.S., *Limited Dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press, 1998.

McFadden, D.L. (1981) Econometric models of probabilistic choice in *Structural Analysis of Discrete Data*, Manski, C.F. and McFadden, D.L. (eds.) MIT Press, Cambridge Massachusetts, 198-271.

Ortuzar, Juan de Dios (1998) A short note on the history of nested logit models, Department of Transport Engineering, Pontificia Universidad Catolica de Chile, Santiago, November.

Quigley, J., "Consumer Choice of Dwelling, Neighborhood, and Public Services," *Regional Science and Urban Economics*, 15, 1985, pp. 41-63.