UNIVERSITY OF LEEDS

This is a repository copy of *Recursive nested extreme value model*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/2067/

Monograph:
Daly, Andrew (2001) Recursive nested extreme value model. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 559

Reuse
See Attached

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

White Rose Research Online
http://eprints.whiterose.ac.uk/

[Institute of Transport Studies](http://www.its.leeds.ac.uk/)
University of Leeds

This is an ITS Working Paper produced and published by the University of Leeds. ITS Working Papers are intended to provide information and encourage discussion on a topic in advance of formal publication. They represent only the views of the authors, and do not necessarily reflect the views or approval of the sponsors.

White Rose Repository URL for this paper:
http://eprints.whiterose.ac.uk/2067/

Published paper
Andrew Daly (2001) *Recursive Nested Extreme Value Model.* Institute of Transport Studies, University of Leeds, Working Paper 559

White Rose Consortium ePrints Repository
eprints@whiterose.ac.uk

ITS Working Paper 559

Recursive Nested Extreme Value Model

Andrew Daly
RAND Europe and ITS Leeds
May 2000, revised April 2001

Contents

1.	Introduction	1
2.	Definition and Basic Properties of RNEV Model	2
3.	Elasticity and Demand Derivatives	5
4.	Estimation Procedure	7
5.	Conclusions	9
	References	10
Appendix 1:	Model Definition	11
Appendix 2:	Proof that RN G is a McFadden GEV Function	13

1. Introduction

The objective of this note is to set out the specification of a quite general member of McFadden's (1981) GEV family of models and to discuss some of the relevant properties of that model.

The key value of McFadden's (1981) result was to establish that a large family of models was consistent with the 'RUM' concept of utility maximisation. The simplest members of that family were the MNL models, for which this consistency had already been known for a few years. The GEV result was also immediately applied to show the consistency of the 'tree logit' model with RUM, a result that was also proved at around the same time by other researchers independently of GEV theory.

Although McFadden indicated further possibilities for the GEV concept, applications of these were slow to materialise, primarily because specialised software was not available and general-purpose software, in particular Gauss, was too slow on the computers of the time to be used for serious modelling. However, in recent years applications have been made: the Paired Comparisons Logit (PCL, Koppelman and Wen, 2000), Cross-Nested Logit (CNL, McFadden, 1981, Vovsha, 1997), Generalised Nested Logit (GNL, Wen and Koppelman, 2000) and Ordered Generalised Extreme Value (OGEV, Small 1987). It is interesting to note that all of these models are of two-level structures: that is, elementary alternatives combine in 'nests' to form composite alternatives (in various ways); choice is then represented as a two stage process of choosing between these nests and then within them. The sole exception to this that has been found in the recent literature is Bhat's (1998) model in which an OGEV is 'grafted' onto the bottom of an MNL model to give a three-level structure. In each case the GEV function has been constructed and shown to satisfy McFadden's requirements, thus showing that the model is consistent with utility maximisation.

In contrast, tree logit models have long been known and even occasionally used in structures with multiple levels. Indefinite structures can be specified and even programmed efficiently (Daly, 1987).

A paper by Dagsvik (1994) indicates that there is a huge variety of GEV models, fitting effectively every possible RUM structure. However, neither Dagsvik nor McFadden indicate how GEV models should be constructed to meet specific requirements.

Recent research appears to be moving away from further exploitation of the GEV family, focussing instead on models of the 'mixed logit' family, which are in many cases easier to construct to meet specific requirements, as McFadden (2000) points out. However, GEV models still offer considerable advantages.

- In GEV models it is much quicker to calculate probabilities and their derivatives, making estimation and application much quicker than with other model forms.[1]
- The existence of the GEV function itself is a considerable advantage for evaluation and use in further modelling.

[1] The form of these models is often called 'closed', neglecting the fact that the evaluation of exp x involves the approximation of an infinite series. Nevertheless, it is true that these evaluations are singly infinite, whereas the evaluation of (e.g.) normal functions requires approximations of doubly infinite series.

– Several of the GEV models can be proved to converge in transport planning applications with conventional assignment procedures (see Prashker and Bekhor, 1999); it is not clear how far these proofs can be extended to more general GEV forms.

A substantial problem with GEV models is that it is necessary to specify the structure in advance in order to be able to work with them. Thus when the structure is at issue this family is not as convenient as those in which the significance of several structural elements can be tested simultaneously. For these reasons it is desirable to be able to write down GEV models with pre-defined properties.

In this note a model is presented which generalises all of the GEV models presented in the literature listed above, as well as generalising a multi-level tree logit model. This is achieved by specifying a recursive nesting structure that allows cross-nesting.

The Recursive Nested Extreme Value (RNEV) model can be specified with a wide range of cross-elasticity properties to fit a wide range of circumstances. The means by which this can be achieved are discussed in Section 3. The fact that this model generalises several other models means that its properties can be used to derive the properties of its special cases.

A further advantage of this models is that there exists an efficient estimation procedure, based on a generalisation – itself useful – of the tree logit estimation method of Daly (1987). This procedure is also discussed. It offers an efficient estimation method, using first and true second derivatives of the likelihood function, for the RNEV model itself and of course for its special cases. These special cases have been estimated by much less efficient means to date, impairing their applicability substantially. The estimation procedure is outlined in Section 4.

2. Definition and Basic Properties of RNEV Model

Given a set E of elementary alternatives, the analyst constructs a structure with an arbitrary number of levels in which alternatives at one level are grouped with arbitrary cross-nesting into nests at the next 'higher' level. At the 'highest' level there is just one nest, which is called the root and labelled r. Choice is then modelled by MNL models at each nest. At each nest, a 'logsum' variable is formed which feeds into choice at the next 'highest' level. A formal definition of the RNEV structure is given in Appendix 1.

Choice in this structure can be modelled by the definition of a function Y for each of the elements of the set D which contains all the nests (including the root) and all the elementary alternatives, as follows

Y_e represents the attractiveness or utility of the choice e
when $e \in E$, i.e. e is elementary;

$Y_d = \Sigma_{k \in N(d)} \lambda_{kd} \, . \, Y_k^{\mu d/\mu k}$ when $d \in D \backslash E$, i.e. d is a nest.

In this definition, N(d) is the set of members of the nest d, whether these are elementary or other nests.

Appendix 1 shows that it is possible to make this definition rigorous, i.e. to define the structure so that it is finite and has no cycles. This in turn implies that the Y's are well-defined.

The parameters λ and μ have to satisfy the constraints that (i) $\lambda, \mu > 0$, (ii) $k \in N(d)$ and $k \notin E$ implies $\mu_k \geq \mu_d$ and (iii) μ_e is constant over the elementary alternatives.

Note that the μ's appear in the model only in the form of ratios. Thus they can all be multiplied by a constant without making any change. This feature will be used to make the normalisation $\mu_r = 1$. *Independently* of the normalisation, it is always possible to redefine

$$Y_e' = (Y_e)^{\mu e}$$

so that μ_e disappears from the model. (The absence of a constraint on the ratio involving μ_e is necessary to make this transformation.) The setting of μ_e and μ_r to 1 will be used to simplify the formulae subsequently. However, for the proofs it is slightly simpler to retain them in the equations.

The function Y_r, i.e. the Y value at the root element, can be viewed as a function of the Y_e for the elementary alternatives, by defining

$$G\ (Y_e, e \in E) = Y_r$$

Appendix 2 proves that G is a GEV function as defined by McFadden (1981).

We can then use the GEV theorem of McFadden to claim that this model is consistent with individual utility maximisation, and that log G is the social surplus function, when the indirect utilities (log Y_e) of the alternatives in E are distributed with the multivariate extreme value distribution.

The choice probabilities of the RNEV model are given by the following formula which is also part of the GEV theorem

$$p_e = \partial(\log Y_r) / \partial(\log Y_e) = (\partial Y_r/\partial Y_e) \,.\, (Y_e / Y_r)$$

At each stage in the recursive definition of the model we have, for $k \in N(d)$

$$\partial Y_d/\partial Y_k = \lambda_{kd} \,.\, (\mu_d/\mu_k) \,.\, Y_k^{(\mu d/\mu k)-1}$$

and so

$$\partial Y_r/\partial Y_e = \Sigma_{\text{all sequences}} \Pi_{k \in N(d)} \lambda_{kd} \,.\, (\mu_d/\mu_k) \,.\, Y_k^{(\mu d/\mu k)-1}$$

where the sum runs over all the sequences that run from the root to alternative e and the product runs over all nests in each sequence. Collecting terms, labelling the set of sequences to elementary alternative e as S_e, and setting μ_e and μ_r to 1, we obtain

$$p_e = \Sigma_{s\in Se} \Pi_{k\in s,\, d=s(k)} \lambda_{kd} \,.\, Y_k^{(\mu d/\mu k)} / Y_d$$

where s(k) denotes the next node after k in the sequence s from e via k to the root.

The λ variables in this model play an interesting role. Effectively, they are alternative-nest-specific constants. If there is also an alternative-specific constant for k, then the λ's will need to be constrained in some way to avoid over-specification. Wen and Koppelman (2000) describe them in the GNL model as 'allocation parameters' and apply the constraint that they must sum to 1; this interpretation is interesting, since it suggests that a part of the alternative is allocated to each sequence. However, it is not clear that the λ's are essential to the specification of the model and the option of setting them all to 1 will often be adopted.

If we define $V_{k(s)} = \log(\lambda_{kd} \,.\, Y_k^{(\mu d/\mu k)})$, where d=s(k), we can write

$$p_e = \Sigma_{s\in Se} [\exp \Sigma_{k\in s} (V_{k(s)} - \log \Sigma_{s(h)=s(k)} \exp V_{h(s)})]$$

and the term in [ ] is just the probability of an alternative in a tree logit model as defined by Daly (1987, 2001b). This formulation also clarifies the role of the λ parameters.

Note that the probability within [ ] is determined by s. Full sequences for alternatives other than e are not relevant to the calculation: such alternatives enter when one of their sequences intersects with s. A useful equivalent formulation of the RNEV model is to see it as an extended TL model, in which each *sequence* from each elementary alternative in the RNEV model is an alternative in the TL model.

From the last equation above we see that the choice probabilities of the alternatives in RNEV are just the sums of tree logit probabilities. From this it is clear that RNEV is a generalisation of (normalised) TL, which represents the RNEV special case in which there is just one sequence S_e for each elementary alternative e. Obviously (therefore) RNEV also generalises MNL.

RNEV also generalises Wen and Koppelman's (2000) GNL, this is the case of RNEV in which the sequences S_e are of maximum length 2. RNEV therefore also generalises PCL[2], CNL, OGEV and the Principles of Differentiation model, all of which are special cases of GNL (see W&K). It also generalises Bhat's (1998) combined OGEV/TL model. Thus any properties possessed by RNEV are also possessed by these special cases. Note that GNL is not a generalisation of TL for more than two levels

I am not aware of *any* GEV model of the McFadden (1981) type that is not a special case of RNEV.

2 Including the Chu variant, the factors $(1 - \sigma)$ can be represented through the λ parameters.

3. Elasticity and Demand Derivatives

Key characteristics of demand models are the 'own' and cross demand differentials, i.e.

$$\partial p_1/\partial V_1 \text{ and } \partial p_1/\partial V_2$$

These functions are of course closely related to the elasticities with respect to a characteristic c, differing from them by a factor $p_1/\beta_c X_c$, where X is the value of the characteristic and β is the differential of V with respect to it. However, both X and p_1 vary in the population, while the presence of β also restricts the applicability of the elasticity, so that the demand differentials are a more stable representation of the sensitivity of the model.

Further, since $\Sigma_k\, p_k = 1$,

$$\partial p_1/\partial V_1 = -\Sigma_{k\neq 1}\, \partial p_k/\partial V_1$$

so that a statement of the cross-derivatives of demand is sufficient to specify the model, the own-derivatives are then implied. The corresponding formulae for elasticities are rather more complicated.

The use of demand differentials has the added advantage in the present context that the differentials for RNEV can be obtained simply by adding the differentials for the corresponding tree logit model. This would of course not be true for elasticities.

For the tree logit model whose probabilities sum to those of the RNEV, the probability of an alternative (i.e. of a sequence in the RNEV) is

$$p_1 = \exp \Sigma_{k\in S1} (V_k - \log \Sigma_{s(h)=s(k)} \exp V_h)$$

To calculate the cross-derivative, we need to identify the lowest common nest to which both alternatives 1 and 2 belong: say x. For choices above x, a change in the utility of 2 has a positive impact on 1, they belong to the same alternative; at and below x the impact is negative because 1 and 2 are competing.

$$\partial p_1/\partial V_2 = p_1 \,.\, \{ \Sigma_{k\in Sx} [(1 - p_{k|s(k)}) \,.\, p_{2|k} \,.\, \mu_{s(k)}] - p_{2|x} .\, \mu_x \}$$

This derivative is non-positive, of course.[3] It is also symmetrical: this is a necessary corollary of the RUM theory.

To focus on the competition between alternatives 1 and 2, it is interesting to eliminate other alternatives from consideration. If we set all the utilities to be –infinity except for those of alternatives 1 and 2, we obtain

---

[3] $\mu_x \geq \mu_k$, for all $k\in sx$, and $p_{2|k} = p_{2|x} \,.\, p_{x|k}$ so it is sufficient to show that

$$\Sigma_{k\in Sx} (1 - p_{k|s(k)}) \,.\, p_{x|k} \leq 1.$$

Because $p_{k|s(k)}.p_{x|k}$ is $p_{x|s(k)}$, the left side of this equation is just

$$\Sigma_{k\in Sx}\, p_{x|k} - p_{x|s(k)} = 1 - p_{x|r},$$

since most of the terms cancel out and $p_{x|x}$ is 1. This is obviously less than 1.

$$\partial p_1/\partial V_2 = -p_1 \cdot p_2 \cdot \mu_x$$

since p_x and all higher choices have probability 1. The negative of this derivative has a maximum at $V_1 = V_2$ which can be defined as the *competitiveness* of the alternatives

$$\psi_{12} = \tfrac{1}{4}\mu_x$$

since $p_1 = p_2 = \tfrac{1}{2}$. The competitiveness between two alternatives in TL (whatever their position in the tree) is proportional to the inverse utility scale μ_x at the point in the tree at which the sequences for those two alternatives meet. For MNL, $\psi_{12} = \tfrac{1}{4}$ in all cases, as it is in TL for alternatives whose sequences meet at the root.

In the RNEV model, the probability of an alternative is given by the sum of a number of TL probabilities and its derivative is thus the sum of the derivatives of those probabilities. The cross-derivative for each other alternative will have a number of impacts, one for each of its sequences. The full cross-derivative is then (when all other alternatives are set to –infinity, and neglecting the λ's)

$$\partial p_1/\partial V_2 = \Sigma_{s1\in S1}\,\Sigma_{s2\in S2}\,\{ -p_{s1} \cdot p_{s2} \cdot \mu_{xs1s2} \}$$

The symmetry of competitiveness measures follows from the symmetry of the demand derivatives that underlie them.

When the utilities of alternatives 1 and 2 are equal, $p_{s1} = p_{s2} = 1/(n_1 + n_2)$, where n_1 and n_2 are the numbers of sequence-alternatives for alternatives 1 and 2 respectively. Under these circumstances, we get a competitiveness[4]

$$\psi_{12} = n_1.n_2 / (n_1+n_2)^2 \,.\, \Sigma_{s1\in S1}\,\Sigma_{s2\in S2}\,\mu_{xs1s2}$$

i.e., the cross-derivative between the utilities of two alternatives is determined by the sum of the utility scales at the points at which all sequences meet. This formula can be used to study the intrinsic properties of RNEV and to compare them with those of other models.

In the PCL we have $n_1 = n_2 = (n-1)$, there are $(n-1)$ sequences for each alternative and all the higher-level μ's are equal, so we get

$$\psi_{12} = \tfrac{1}{4}\,\{ (n^2 - 2n) + \mu \}$$

For the GNL, there is little simplification relative to the RNEV formula and we get

$$\psi_{12} = n_1.n_2 / (n_1+n_2)^2 \,.\, \{ (n_1 + n_2 - n_{12}) + \Sigma_{1,2\in N(x)}\,\mu_x \}$$

where n_{12} is the number of nests that contain *both* alternatives 1 and 2, i.e. the size of the set $\{ x \mid 1,2\in N(x) \}$ over which the sum is calculated.

[4] This value is not necessarily the maximum of the demand derivative in cross-nested models, unless all the λ's are equal.

An alternative comparison is that with a probit model. For a multinomial probit model, it is easy to calculate

$$\psi_{12} = 1 / \sigma_{12}\sqrt{(2\pi)}$$

where σ_{12} is the standard deviation of the utility difference between alternatives 1 and 2[5]. In this case, there is a simple one-to-one relationship between the competitiveness ψ and the variance of utility differences, so that a statement of the matrix of competitivenesses specifies the probit model entirely, up to the set of 'indistinguishable' models (Daly, 2001a). Of course, before starting on such an exercise, one would be well advised to check that a model with the required properties might exist.

An interesting issue for further research is the extent to which it is possible to design GEV models, for example RNEV models, to match any given set of competitiveness figures. As in the case of probit models, it would be useful to check in advance that the specified set of competitivenesses might exist.

In a TL model, given an existing structure, it is always possible to introduce a new alternative that has *any* required competitiveness with *any* existing alternative. However, the competitiveness with other alternatives is then fixed. A TL model for n alternatives has a maximum of n–1 nests (including the root), so a severe restriction on competitiveness must be expected.

In a GNL model, there is a maximum of n(n–1)/2 nests, each with its own parameter, so in principle the number of degrees of freedom is adequate to represent any desired competitiveness pattern. However, if two alternatives are required each to have specific patterns of interaction with other alternatives, then they have a minimum level of competitiveness with each other, because all the other nests are connected to the root. Therefore it is difficult to introduce a pattern of low and high competitiveness into the model.

In an RNEV model, greater freedom exists and it is possible to imagine complex patterns of interaction between subsets of alternatives, which then have complex patterns of interaction with each other. A greater range of high and low competitiveness can be specified. But the full working out of the possibilities of the model is the subject for future research.

4. Estimation Procedure for RNEV

It has been shown that RNEV predicts choice as the sum of a number of tree logit probabilities. This specification of the model can be extended to allow the estimation of RNEV models as tree logit models.

Consider a tree logit model in which the elementary and composite alternatives are defined by the *sequences* leading to elementary and composite alternatives in an

[5] Note that if we set $\sigma = \pi / \sqrt{3}$, so that the normal distribution has the same variance as would be obtained from a standard logistic distribution, we obtain $\psi = 0.220..$, a rather lower value than for the 'almost indistinguishable' (Daly, 2001a) MNL model (0.25).

RNEV model. At each stage in the two models, the conditional choice probabilities are the same, and the sequences define sets of ancestors as required for tree logit.

It would thus be possible to estimate RNEV as a moderate extension of the capabilities of a tree logit estimation program. The extension, which would require the acceptance that choice could be an unknown member of a specified subset of alternatives, would also be useful in other contexts[6].

For a general choice model, the kernel of the likelihood function for a choice falling within a set is given by

$$\log L \;=\; \Sigma_{obs}\, l_{obs} \;=\; \Sigma_{obs} \log p_{obs} \;=\; \Sigma_{obs}\; \log \{\, \Sigma_{k\in c(obs)}\, p_{k,obs} \,\}$$

where c(obs) is the set of alternatives containing the choice for a given observation and $p_{k,obs}$ is the probability of choice k for that observation.

The contribution of each observation to the first and second differentials of this function are as follows (using subscripts to indicate differentiation with respect to the unknown parameters)

$$l_i \;=\; (1/p)\,.\,\Sigma_{k\in c}\, p_{ki}$$

$$l_{ij} \;=\; (1/p)\,.\,\Sigma_{k\in c}\, p_{kij} \;-\; (1/p^2)\,.\,\Sigma_{k\in c}\, p_{ki}.\, \Sigma_{k\in c}\, p_{kj} \;=\; (1/p)\,.\,\Sigma_{k\in c}\, p_{kij} \;-\; l_i\, l_j$$

In the present case, these probabilities are those of a tree logit model and are given by[7]

$$\log p_k \;=\; \Sigma_{a\in S(k)}\, (\, V_a - \log \Sigma_{s(b)=s(a)} \exp V_b \,)$$

where S(k) is the sequence (set of ancestors) of k (including k but excluding the root) induced by the tree function s.

Hence

$$p_{ki} \;=\; p_k\,.\,\Sigma_{a\in S(k)}\, (\, V_{ai} - V^*_{s(a)i} \,)$$

where V_{ai} is the differential of the (indirect) utility function of alternative a with respect to unknown parameter i;
$V^*_{si} = \Sigma_{s(b)=s}\, p_{b|s}\, V_{bi}$; and
$p_{b|s}$ is the conditional probability of b, a member of nest s, given that the nest is chosen.

and

[6] Current software can accommodate choice within a subset of alternatives but only if that subset forms a nest in the tree structure. This feature has proved useful in several practical contexts but has slightly limited the model specification search.

[7] The following equations apply to both normalised and non-normalised logit models, the differences can be expressed within the V functions alone, see Daly (2001b).

$$
\begin{aligned}
p_{kij} &= p_{kj} \,.\, \Sigma_{a\in S(k)} (V_{ai} - V^*_{s(a)i}) \\
&\quad + p_k \,.\, \Sigma_{a\in S(k)} (V_{aij} - V^*_{s(a)ij} - \Sigma_{s(b)=s(a)} p_{b|s(a),j} V_{bj})) \\
&= p_{ki} \,.\, p_{kj} / p_k \\
&\quad + p_k \,.\, \Sigma_{a\in S(k)} (V_{aij} - V^*_{s(a)ij} - \Sigma_{s(b)=s(a)} p_{b|s(a)}.V_{bi}.V_{bj} + V^*_{s(a)i}.V^*_{s(a)j})
\end{aligned}
$$

where V_{aij} is the differential of the (indirect) utility function of alternative a with respect to unknown parameters i and j; and
$V^*_{sij} = \Sigma_{s(b)=s} p_{b|s} V_{bij}$.

Often, V_{aij} will be zero for elementary alternatives as V is often linear in the unknown parameters. However, even in this case $[l_{ij}]$ may be non-definite because V_{aij} is not zero for composite alternatives. These formulae are relatively simple to calculate and many components are already available in the existing software.

The advantage of using true second derivatives of the likelihood function has been established over a number of years. A typical alternative approach is that used by Bhat (1998), who programs the likelihood function and its first derivatives of his OGEV-MNL combined model, then uses the capabilities of the Gauss software to find the optimum. The latter approach is obviously less efficient in computer time – both because of the use of general-purpose rather than specialised software and because the successive steps in the optimisation use approximate rather than exact second derivatives – and does not give the correct values of the error measures of the estimated parameters at the optimum. The Gauss approach is suited to ground-breaking research rather than production work.

An algorithm exploiting the true second derivatives must be prepared to cope with situations in which the matrix is not negative definite, but procedures for dealing with this situation are well established.

5. Conclusions

A model structure, the Recursive Nested EV model, has been defined which is believed to be novel and which generalises all known EV models. This structure is consistent with the GEV theory.

Elasticity, demand derivative and competitiveness formulae can be derived for the RNEV model. They give greater flexibility than previous models but it is not yet known to what extent a pre-specified structure can be implemented.

The RNEV model can be estimated relatively easily by viewing it as the sum of a number of intertwined tree logit models. Of course this estimation procedure would also apply for the special cases of the model. The first and second derivatives of the likelihood function can be written down and would not require great difficulty to program. The program would also cover a useful practical case when the choice in a tree logit (or other RNEV special case) model was known only as a member of a subset.

References

Ben-Akiva, M. and Lerman, S. R. (1985) Discrete Choice Analysis, MIT Press.

Bhat C. R. (1998) Analysis of Travel Mode and Departure Time Choice for Urban Shopping Trips, Transportation Research B, **32**, pp. 361-371.

Dagsvik, J. (1994) Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Alternatives, Econometrica, **62**, pp. 1179-1205.

Daly, A. J. (1987) Estimating 'Tree' Logit Models, *Transportation Research*, 21B, pp 251- 267.

Daly, A. J. (2001a) Independence, Homoskedasticity and Existence in Random Utility Models, ITS Working Paper 557, University of Leeds.

Daly, A. J. (2001b) Comments on a paper of Koppelman and Wen, to be published in Transportation Research.

Koppelman, F. S. and Wen, C.-H. (2000) The paired combinatorial logit model: properties, estimation and application, Transportation Research B, **34**, pp. 75-89.

McFadden, D. L. (1981) Econometric Models of Probabilistic Choice, in McFadden and Manski (eds.) Structural Analysis of Discrete Data with Econometric Applications, MIT Press.

McFadden, D. L. (2000) Disaggregate Behavioural Travel Demand's RUM Side: A 30-year retrospective, presented at International Association for Travel Behaviour Conference, Gold Coast, Queensland.

Prashker, J. N. and Bekhor, S. (2000) Stochastic User-Equilibrium Formulations for Extended-Logit Assignment Models, *Transportation Research Record 1676*, paper no. 99-0645.

Small, K. A. (1987) A discrete choice model for ordered alternatives, Econometrica **55** (2), pp. 409-424.

Vovsha, P. (1997) Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record*, 1607, pp6-15.

C.-H. Wen and F. S. Koppelman, The Generalized Nested Logit Model, TRB, 2000.

Appendix 1: Model Definition

Consider a model of discrete choice, i.e. giving the probability of choice of a single alternative from a finite set E. Let D be a finite superset of E, let M be the set of non-empty subsets of D and let N be a single-root non-circular function

$$N: \quad D \backslash E \rightarrow M$$

The set N(d) is called the nest of d.

'Non-circular' in this context means that there exist no cycles

$$d, d' \in N(d), d'' \in N(d'), \ldots d \quad \text{for any member d of D.}$$

In particular $d \notin N(d)$. If N is non-circular there exists an indexing function I on D (i.e. a mapping from D to the set of positive integers) such that

$$\text{if } d' \in N(d) \text{ then } I(d') < I(d).$$

Conversely, the existence of such a mapping implies that N is non-circular.

This structure also implies that the sequences d, $d' \in N(d)$, .. always end with an element d* such that $N(d^*) \subset E$.

There exists an element r of D which has the highest value of I; r will be called the 'root'. It clearly cannot be a member of the nest of any other member of D.

'Single-root' in this context means that every member of D except the root is represented at least once in the nests defined by N. This property, together with non-circularity, implies that for any member d* of D that there exists at least one sequence

$$r, d \in N(r), d' \in N(d), .. d^*$$

In particular this applies for members of E. Let S_e be the set of sequences from the root to $e \in E$, defining each $s \in S_e$ to include e but exclude r. In each sequence, each lower-level alternative k has a unique higher-level alternative d such that $k \in N(d)$; we define this higher-level alternative to be s(k). Note that s is a tree function and the sequence S_e itself forms the set of 'ancestors' of each elementary alternative e as defined by Daly (1987).

The idea here is that choice is represented as successive choices from subsets. So when d has been chosen, the next choice is one of the elements of the 'nest' N(d) and so on until we arrive at an elementary alternative. There is no insistence that nests should not overlap, so this structure is a generalisation of strict 'tree' models. The fact that choices can be defined over an indefinite series of nests will mean that the structure can generalise the Wen and Koppelman (2000) Generalised Nested Logit, which is restricted to two levels.

The model is called 'nested' and the composite nodes D\E could be defined to be 'nests' containing the elementary alternatives in any of whose sequences they fall but (a) there may be more than one nest with the same set of alternatives and (b) this is not essential to the definition: D\E can be any arbitrary construct.

The function I is not central to the definition, it is set up only to guarantee non-circularity and it also plays a useful role in the proofs.

Given this non-circular nesting and a positive function Y for elementary alternatives, we can define Y for the remaining elements of D recursively

$$Y_d = \Sigma_{k \in N(d)} \lambda_{kd} . Y_k^{\mu d/\mu k}$$

subject to the constraints that (i) $\lambda, \mu > 0$, (ii) $k \in N(d)$ and $k \notin E$ implies $\mu_k \geq \mu_d$ and (iii) μ_e is constant over the elementary alternatives.

Note that the Y's are also well defined because of the non-circularity of N (a proof can be given easily by induction using I).

Y_r, the Y function for the root, can be viewed as a function of the Y's for the elementary alternatives. Define

$$G = Y_r (Y_e, e \in E).$$

This function is proved to be a McFadden GEV function in Appendix 2.

Appendix 2: Proof that RN G is a McFadden GEV Function

The function G is defined as in Appendix 1. The required properties for it to be GEV are given by Ben-Akiva and Lerman (1985) as follows.

GEV1 G is non-negative: this is obvious from the form of Y, and the constraints on Y_e and λ.

GEV2 G is homogenous. It can easily be calculated that

$$G(\alpha Y_1, \alpha Y_2, ..) = \alpha^{\mu r/\mu e} G(Y_1, Y_2, ..)$$

This follows from the non-circularity of N, the Y definitions and the fact that μ_e is constant. Thus G is homogenous of degree μ_r/μ_e which is positive and therefore satisfies the extended Ben-Akiva/Francois definition of GEV.

GEV3 G goes to infinity with any of the Y_e's. This follows from the single-root definition of N, i.e. all elementary alternatives 'feed in' to the same root, and the constraints on λ and μ.

GEV4 The mixed partial derivatives of G with respect to Y_e are continuous, with non-positive even and non-negative odd mixed partial derivatives. It is clear that all the derivatives exist and are continuous, from the form of Y and the restrictions on μ. Proof of this property is given below.

McFadden (1981) gives a further condition for GEV, seemingly not required by Ben-Akiva and Lerman, which appears to state only that the addition of infinitely bad alternatives (Y=0) to the choice set does not alter the choice. This appears to be necessary to ensure the symmetry of partial derivatives. I have assumed that the Ben-Akiva-Lerman conditions are sufficient to define a GEV function.

GEV4 is proved by first establishing a Lemma. For convenience of notation, we define $\theta_{dk} = \mu_d/\mu_k$ when $k \in N(d)$ (the constraints on μ imply that $0 < \theta_{dk} \leq 1$ when $k \notin E$) so we can write

$$Y_d = \Sigma_{k \in N(d)} \lambda_{kd} . Y_k^{\theta dk}$$

<u>Lemma</u>: For distinct Y_1, .. Y_m, and for any $d \in D\backslash E$,

$$\partial^r Y_d / \partial Y_1 ... \partial Y_m = \Sigma_{k \in N(d)} \lambda_{kd} \Sigma_{i=1,m} \pi_{\theta dk,i} . Y_k^{\theta dk-i} . S_{imk}$$

where $\pi_{\theta,i} = \theta . (\theta - 1) . (\theta - 2) .. (\theta - i + 1)$, i.e. $\pi_{\theta,1} = \theta$

and S_{imk} is the sum of all different products of i partial derivatives of any order of Y_k with respect to distinct Y_j, j =1, .. m, such that each Y_j appears exactly once; the sum of the orders of the partial derivatives in each product is then m.

<u>Proof</u>
Simply differentiating Y_d we obtain

$$\partial Y_d / \partial Y_1 = \Sigma_{k\in N(d)} \lambda_{kd} . \theta_{dk}. Y_k^{\theta dk-1} . \partial Y_k / \partial Y_1$$

$$\partial^2 Y_d / \partial Y_1 \partial Y_2 = \Sigma_{k\in N(d)} \lambda_{kd} . \{ \theta_{dk}. Y_k^{\theta dk-1} . \partial^2 Y_k / \partial Y_1 \partial Y_2 + \theta_{dk} (\theta_{dk} - 1) Y_k^{\theta dk-2} . \partial Y_k / \partial Y_1 . \partial Y_k / \partial Y_2 \}$$

and these have the form required. It is thus sufficient to show that the m^{th} derivative implies the $m+1^{th}$ to prove the entire Lemma by induction on the order of the partial derivative. Differentiating the inductive hypothesis, we obtain

$$\partial^{r+1} Y_d / \partial Y_1 ... \partial Y_{m+1} = \Sigma_{k\in N(d)} \lambda_{kd} \Sigma_{i=1,m} \pi_{dk,i} . \{ (\theta_{dk} - i). Y_k^{\theta dk-i-1} . \partial Y_k / \partial Y_{m+1} . S_{imk} + Y_k^{\theta dk-i} . \partial S_{imk} / \partial Y_{m+1} \}$$

and we can always express

$$S_{i(m+1)k} = \partial Y_k / \partial Y_{m+1} . S_{(i-1)mk} + \partial S_{imk} / \partial Y_{m+1}$$

the former term giving all the partial derivative products **including** $\partial Y_k/\partial Y_{m+1}$, the latter term giving all the products excluding that first-order differential. Note that S_{1mk} has only the single m^{th} order differential with respect to all the Y's (when S_{1mk} is expressed as in the last equation above the first term does not exist), while S_{mmk} is just the product of the m first partial derivatives (the second term in the last equation does not exist). For all other values of i, appropriate contributions are made to $S_{i(m+1)k}$ by the first term in { } brackets for (i–1) and the second term in { } for i.

Thus the Lemma is proved.

Note that the second and higher mixed derivatives of alternatives $Y_{s(e)}$ directly above elementary alternatives, with respect to the Y_e functions of those elementary alternatives e, have a particularly simple form. The first derivative with respect to Y_e contains *only* a term in the same Y_e, so that all the second and higher mixed derivatives are zero, even if only one of the components relates to an elementary alternative in the nest of s(e).

GEV4 and thus the entire result is then proved by induction on the index I (defined in Appendix 1) of each alternative, as follows.

The alternative with the lowest value of I is certainly an elementary alternative and for such alternatives the partial derivative with respect to itself is 1 and with respect to all other Y's is 0. Both these possible values are non-negative. The higher-order derivatives are all zero which of course satisfies the requirement.

Suppose that GEV4 holds for values of I less than that of alternative d and consider the sign of the partial derivatives of Y_d. Either d is elementary, in which case there is no problem, or it has a nest. All the alternatives in the nest have partial derivatives satisfying GEV4, by the inductive hypothesis, because their I value is less than that of d.

Each product in S_{imk} has the sum of its orders equal to m, thus the number of odd-order derivatives in each product is equal to m modulo 2. Since there are i components in each product, the number of even-order derivatives is equal to i-m modulo 2. Since the even-order derivatives are non-positive (by inductive hypothesis), each product in S_{imk} (and hence S_{imk} itself) is non-positive if i-m is odd, non-negative if i-m is even.

$\pi_{\theta dk,i}$ is non-negative if i is odd, non-positive if i is even, when $k \notin E$, since $0 < \theta_{dk} \leq 1$. The product of $\pi_{\theta dk,i}$ and S_{imk} is then non-positive if m is even, non-negative if m is odd. When $k \in E$, the derivatives have the simple form as noted above, so that S_{1mk} is positive and S_{imk} is zero for i>1. Thus the partial derivatives for alternative d also satisfy the hypothesis.

By induction this proof can be extended to cover all the alternatives and in particular the root r, thus establishing GEV4.