The Network GEV model

Michel Bierlaire, EPFL

Conference paper STRC 2002

The Network GEV model

Michel Bierlaire, Dpt of Mathematics, EPFL, Lausanne

Phone:	+41.21.693.25.37
Fax:	+41.21.693.55.70
Email:	michel.bierlaire@epfl.ch

Abstract

Generalized Extreme Value models provide an interesting theoretical framework to develop closed form random utility models. Recently, several members of this family have been proposed in the literature. These models, although different, exhibit several similarities. Each of them must be proven to belong to the GEV family, and is difficult to esitmate. In this paper, we propose the Network GEV model, a new modeling approach providing an intuitive way to generate a wide class of concrete Generalized Extreme Value (GEV) models. Using this approach, the analyst needs only to design a network structure capturing the underrlying correlation structure of the considered application. If the network complies with some simple conditions, we prove that the associated Network GEV model is indeed a GEV model and, therefore, complies with random utility theory. The Multinomial Logit, the Nested Logit and the Cross-Nested Logit models are members of the Network GEV models class. The recent GenL model, combining choice set generation and choice model, is also a Network GEV model.

Keywords: GEV model, random utility, transportation demand

Swiss Transport Research Conference – STRC 2002 – Monte Verita

1 Introduction

Discrete choice models play a major role in many fields involving a human dimension, including econometry, marketing research and transportation demand analysis. Their nice and strong theoretical properties, and their flexibility to capture various situations, provide a vast topic of interest for both researchers and practitioners, that has (by far) not been totally exploited yet. The theory on Generalized Extreme Value (GEV) models has been introduced by McFadden (1978). It provides a tremendous potential, as it defines a whole family of models, consistent with random utility theory. It appears that only a few members of this family have been exploited so far, the Multinomial Logit model and the Nested Logit model being the most popular (Ben-Akiva and Lerman, 1985). Recent research on the Cross-Nested logit model (Small, 1987, Vovsha, 1997, Vovsha and Bekhor, 1998, Ben-Akiva and Bierlaire, 1999, Papola, 2000, Bierlaire, 2001a, Wen and Koppelman, 2001, Swait, 2001) has slightly extended the number of GEV models used in practice.

The most general GEV model published thus far is probably the Recursive Nested Extreme Value (RNEV) model, proposed by Daly (2001). It is an elegant generalization of the Cross-Nested logit model, where multiple layers of nests are allowed. RNEL is designed to be easily estimated, as it requires moderate extensions to nested logit estimation packages like ALOGIT (Daly, 1987) or HieLoW (Bierlaire, 1995, Bierlaire and Vandevyvere, 1995).

In this paper, we propose a new modeling approach, providing an intuitive way of generating a wide class of concrete GEV models. The idea is an extension of the use of trees to represent Nested Logit models (Ben-Akiva and Lerman, 1985, Daly, 1987). Here, we base the model definition on a network representation. The advantages of this approach are the following.

- We formally prove that any model based on a network representation complying with some simple properties is indeed a GEV model. Therefore, consistency with random utility theory is guaranteed.
- A network representation allows to intuitively capture complex correlation structures of actual modeling situations. This feature, intensively exploited with trees for the Nested Logit models in the literature, can now be extended to a wide class of GEV models.
- The recursive definition of the model, based on the network structure, greatly simplifies its formulation.

The main objective of the paper is to provide a general theorical result, such that the development of new GEV models will be easier in the future. Indeed, in addition to the intuitive approach due to the network structure, any instance of the Network GEV model is proven to be a GEV model and therefore, no more theoretical justification is required for such models.

The Network GEV model is defined in Section 3, where we prove that it is indeed a GEV model. The proof is based on technical lemmas developed in Appendix A. In Section 4, we provide a couple of concrete examples of Network GEV models, and discuss some practical issues.

2 The GEV model

The Generalized Extreme Value (GEV) model has been derived from the random utility model by McFadden (1978). This general model consists of a large family of models that include the Multinomial Logit, the Nested Logit and the Cross-Nested Logit models. The probability of choosing alternative i within the choice set C of a given choice maker is

$$P(i|C) = \frac{y_i \frac{\partial G}{\partial y_i}(y_1, \ldots, y_J)}{\mu G(y_1, \ldots, y_J)} \tag{1}$$

where J is the number of available alternatives, $y_i = e^{V_i}$, V_i is the deterministic part of the utility function associated to alternative i, and G is a non-negative differentiable function defined on $\mathbb{R}_+^J$ with the following properties:

1. G is homogeneous of degree $\mu > 0$, that is $G(\alpha y) = \alpha^\mu G(y)$,
2. $\lim_{y_i \to +\infty} G(y_1, \ldots, y_i, \ldots, y_J) = +\infty$, for each $i = 1, \ldots, J$,
3. the kth partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even that is, for any distinct indices $i_1, \ldots, i_k \in \{1, \ldots, J\}$, we have

$$(-1)^k \frac{\partial^k G}{\partial x_{i_1} \ldots \partial x_{i_k}}(x) \leq 0, \; \forall x \in \mathbb{R}_+^J. \tag{2}$$

Note that the homogeneity of G and Euler's theorem give

$$P(i|C) = \frac{e^{V_i + \ln G_i(\ldots)}}{\sum_{j=1}^J e^{V_j + \ln G_j(\ldots)}}, \tag{3}$$

where $G_i = \frac{\partial G}{\partial y_i}$.

3 The Network GEV model

Let (V, E) be a directed graph, where V is the set of vertices and E the set of edges. Each edge (i, j) is associated with a non-negative parameter $\alpha_{(i,j)} \geq 0$, so that the directed graph is a network. The network has the following properties:

1. It does not contain any circuit.

2. It has one special node with no predecessor, called the *root*, and denoted by v_0.

3. It has J special nodes with no successor, called the *alternatives*, and denoted by $v_1, \ldots, v_J$.

4. For each node v_i in the network, there exists at least a path (v_{i_0}, v_{i_1}), (v_{i_1}, v_{i_2}), …, $(v_{i_{P-1}}, v_{i_P})$ connecting $v_0 = v_{i_0}$ and $v_i = v_{i_P}$ such that

$$\prod_{k=1}^{P} \alpha_{(i_{k-1}, i_k)} > 0, \tag{4}$$

   that is all parameters on the path are non-zero.

We associate with each node v_i of the network

- a set $I_i \subseteq \{1, \ldots, J\}$ of J_i relevant alternatives,

- a homogeneous function $G^i : \mathbb{R}^{J_i} \longrightarrow \mathbb{R}$, and

- an homogeneity parameter μ_i.

We define $I_i = \{i\}$ for the nodes representing the alternatives, that is for $i = 1, \ldots, J$, and $I_i = \bigcup_{j \in \mathrm{succ}(i)} I_j$ for all others. Note that we can deduce from network property 4 that $I_0 = \{1, \ldots, J\}$. The homogeneous functions are defined as follows.

$$G^i : \mathbb{R} \longrightarrow \mathbb{R} : G^i(x_i) = x_i^{\mu_i} \quad i = 1, \ldots, J, \tag{5}$$

and

$$G^i : \mathbb{R}^{J_i} \longrightarrow \mathbb{R} : G^i(x) = \sum_{j \in \text{succ}(i)} \alpha_{(i,j)} G^j(x)^{\frac{\mu_i}{\mu_j}}. \tag{6}$$

If $\mu_i \leq \mu_j$ for each edge (i, j) such that $\alpha_{(i,j)} \neq 0$, the function G^0 associated with the root node, and entirely defined by the network structure, is a GEV generating function. Indeed, $G^1, \ldots, G^J$ defined by (5) trivially verify the GEV conditions. The G^i functions (6) associated with all other nodes, including the root, also verify the GEV conditions, as proven by Theorem 1.

Theorem 1 *If $G^i : \mathbb{R}^J \longrightarrow \mathbb{R}$, $i = 1, \ldots, p$ are GEV generating functions with homogeneity factor μ_i, than the function $G : \mathbb{R}^J \longrightarrow \mathbb{R}$, defined by*

$$G(x) = \sum_{i=1}^{p} \alpha_i G^i(x)^{\frac{\mu}{\mu_i}} \tag{7}$$

is also a GEV generating function with homogeneity factor μ if the following conditions are verified:

1. $\alpha_i \geq 0$, $i = 1, \ldots, p$,

2. $\sum_{i=1}^{p} \alpha_i > 0$

3. $\mu > 0$,

4. $\mu_i > 0$, $i = 1, \ldots, p$,

5. $\mu \leq \mu_i$, $i = 1, \ldots, p$.

Proof.

1. G is obviously non negative, if $x \in \mathbb{R}^J_+$.

2. G is homogeneous of degree μ. Indeed,

$$\begin{array}{rcl} G(\beta x) & = & \sum_{i=1}^p \alpha_i G^i(\beta x)^{\frac{\mu}{\mu_i}} \\ & = & \sum_{i=1}^p \alpha_i (\beta^{\mu_i} G^i(x))^{\frac{\mu}{\mu_i}} \text{ (as } G^i \text{ is homogeneous of degree } \mu_i), \\ & = & \sum_{i=1}^p \alpha_i \beta^{\mu} G^i(x)^{\frac{\mu}{\mu_i}} \\ & = & \beta^{\mu} \sum_{i=1}^p \alpha_i G^i(x)^{\frac{\mu}{\mu_i}} \\ & = & \beta^{\mu} G(x). \end{array} \tag{8}$$

3. We have

$$\begin{array}{rcl} \lim_{x_k \to \infty} G(x) & = & \lim_{x_k \to \infty} \sum_{i=1}^p \alpha_i G^i(x)^{\frac{\mu}{\mu_i}} \\ & = & \sum_{i=1}^p \alpha_i \lim_{x_k \to \infty} G^i(x)^{\frac{\mu}{\mu_i}}. \end{array} \tag{9}$$

From assumption 3, and because each G^i verifies the GEV assumptions, we have $\lim_{x_k \to \infty} G^i(x)^{\frac{\mu}{\mu_i}} = \infty$. The limit property holds because not all coefficients α_i can be zero, from assumption 1.

4. Without loss of generality, we consider the kth derivative of $G(x)$ with respect to $x_1, \ldots, x_k$, that is

$$\frac{\partial^k G(x)}{\partial x_1 \ldots \partial x_k} = \sum_{i=1}^p \alpha_i \frac{\partial^k G^i(x)^{\frac{\mu}{\mu_i}}}{\partial x_1 \ldots \partial x_k}. \tag{10}$$

The sign of (10) is entirely determined by the sign of $\frac{\partial^k G^i(x)^{\frac{\mu}{\mu_i}}}{\partial x_1 \ldots \partial x_k}$, as the coefficients α_i are non-negative. Using Lemma 2 with $\beta = \frac{\mu}{\mu_i}$, we have

$$\frac{\partial^k}{\partial x_1 \ldots \partial x_k} G(x)^{\frac{\mu}{\mu_i}} = \sum_{P \in \mathcal{P}_k} S^p \prod_{R \in P} D_R, \tag{11}$$

		(1)	(2)	(3)
k	p	S^p	$\prod_{R\in P} D_R$	$S^p \prod_{R\in P} D_R$
even	even	≤ 0	≥ 0	≤ 0
even	odd	≥ 0	≤ 0	≤ 0
odd	even	≤ 0	≤ 0	≥ 0
odd	odd	≥ 0	≥ 0	≥ 0

Table 1: Sign of $S^p \prod_{R\in P} D_R$

where

$$S^p(x) = \prod_{i=0}^{p-1} (\frac{\mu}{\mu_i} - i) G(x)^{\frac{\mu}{\mu_i} - p}, \tag{12}$$

and

$$D_R = \frac{\partial^r G(x)}{(\partial x_i)_{i\in R}}. \tag{13}$$

Consequently, we need to analyze the sign of the right-hand side in (11). Interestingly, for a given k, all terms of this sum have the same sign. If we consider an arbitrary partition $P \in \mathcal{P}_k$, composed of p subsets, the sign of the term for all possible cases is analyzed in Table 1. Column (1) is obtained from Lemma 3, which can be applied because $0 < \mu/\mu_i \leq 1$. Column (2) is obtained from Table 2 in Lemma 4. And column (3) is simply the product of (1) and (2).

Therefore, the sign of (11) and consequently of (10) is non-positive if k is even, and non-negative if k is odd.

□

4 Examples and practical issues

We present here some existing models that are special cases of the Network GEV model. First, the tree representation of Nested Logit model is a Network GEV model, where the network is obviously the tree, and the α parameters associated with the edges are all 1. The Multinomial Logit Model being a special case of the Nested Logit, it is also a Network GEV model. The Cross-Nested model is a GEV model generated by

$$G(x_1, \ldots, x_J) = \sum_m \left(\sum_{j \in C} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}. \tag{14}$$

It is a special case of the Network GEV model, where the network is composed of a root v_0, a list of nodes for the alternatives $v_1, \ldots, v_J$ and a list of nodes $w_1, \ldots, w_M$ for the nests. There is an edge between the root and each nest m, with a parameter α_{0m}, and an edge between each nest m and each alternative i, with a parameter α_{mi}. The μ_i associated to each alternative are set to 1, and we obtain a GEV model generated by the following function:

$$G^0(x_1, \ldots, x_J) = \sum_m \alpha_{0m} \left(\sum_{i=1}^{J} \alpha_{mi} x_i^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \tag{15}$$

Let $\bar{\alpha}_{im} = \alpha_{0m}^{\frac{\mu_m}{\mu}} \alpha_{mi}$, and we obtain the generating function (14) of the Cross-Nested Logit model. As an example, we consider the simple network represented in Figure 1. We have

$$G^1 = x_1^{\mu_1}, G^2 = x_2^{\mu_2}, G^3 = x_3^{\mu_3}.$$

and

$$
\begin{aligned}
G^4 &= \alpha_{41}(G^1)^{\frac{\mu_4}{\mu_1}} + \alpha_{42}(G^2)^{\frac{\mu_4}{\mu_2}} + \alpha_{43}(G^3)^{\frac{\mu_4}{\mu_3}} \\
&= \alpha_{41}x_1^{\mu_4} + \alpha_{42}x_2^{\mu_4} + \alpha_{43}x_3^{\mu_4} \\
G^5 &= \alpha_{51}(G^1)^{\frac{\mu_5}{\mu_1}} + \alpha_{52}(G^2)^{\frac{\mu_5}{\mu_2}} + \alpha_{53}(G^3)^{\frac{\mu_5}{\mu_3}} \\
&= \alpha_{51}x_1^{\mu_5} + \alpha_{52}x_2^{\mu_5} + \alpha_{53}x_3^{\mu_5}
\end{aligned}
$$

Finally,

$$
\begin{aligned}
G^0 &= \alpha_{04}(G^4)^{\frac{\mu_0}{\mu_4}} + \alpha_{05}(G^5)^{\frac{\mu_0}{\mu_5}} \\
&= \alpha_{04}\left(\alpha_{41}x_1^{\mu_4} + \alpha_{42}x_2^{\mu_4} + \alpha_{43}x_3^{\mu_4}\right)^{\frac{\mu_0}{\mu_4}} + \\
&\quad\ \alpha_{05}\left(\alpha_{51}x_1^{\mu_5} + \alpha_{52}x_2^{\mu_5} + \alpha_{53}x_3^{\mu_5}\right)^{\frac{\mu_0}{\mu_5}} \\
&= \left[\alpha_{04}^{\frac{\mu_4}{\mu_0}}\alpha_{41}x_1^{\mu_4} + \alpha_{04}^{\frac{\mu_4}{\mu_0}}\alpha_{42}x_2^{\mu_4} + \alpha_{04}^{\frac{\mu_4}{\mu_0}}\alpha_{43}x_3^{\mu_4}\right]^{\frac{\mu_0}{\mu_4}} + \\
&\quad\ \left[\alpha_{05}^{\frac{\mu_5}{\mu_0}}\alpha_{51}x_1^{\mu_5} + \alpha_{05}^{\frac{\mu_5}{\mu_0}}\alpha_{52}x_2^{\mu_5} + \alpha_{05}^{\frac{\mu_5}{\mu_0}}\alpha_{53}x_3^{\mu_5}\right]^{\frac{\mu_0}{\mu_5}} \\
&= \left[\left(\alpha_{04}^{\frac{1}{\mu_0}}\alpha_{41}^{\frac{1}{\mu_4}}x_1\right)^{\mu_4} + \left(\alpha_{04}^{\frac{1}{\mu_0}}\alpha_{42}^{\frac{1}{\mu_4}}x_2\right)^{\mu_4} + \left(\alpha_{04}^{\frac{1}{\mu_0}}\alpha_{43}^{\frac{1}{\mu_4}}x_3\right)^{\mu_4}\right]^{\frac{\mu_0}{\mu_4}} + \\
&\quad\ \left[\left(\alpha_{05}^{\frac{1}{\mu_0}}\alpha_{51}^{\frac{1}{\mu_5}}x_1\right)^{\mu_5} + \left(\alpha_{05}^{\frac{1}{\mu_0}}\alpha_{52}^{\frac{1}{\mu_5}}x_2\right)^{\mu_5} + \left(\alpha_{05}^{\frac{1}{\mu_0}}\alpha_{53}^{\frac{1}{\mu_5}}x_3\right)^{\mu_5}\right]^{\frac{\mu_0}{\mu_5}}
\end{aligned}
$$

Note that removing an edge of the network amounts to set the associated parameter to 0.

The issue of model estimability must be analyzed. First, the value of the parameters $\mu_1, \ldots, \mu_J$ is irrelevant for the model. Therefore, these can be set to 1 without loss of generality. The other parameters μ are relevant only in terms of their ratio, like for Nested Logit models. Therefore, a normalization (from the top or the bottom) is required. Finally, not all parameters α associated with the edges of the network can be identified. Actually, from a random utility model viewpoint, the meaningful parameters are associated with paths between the root node and the alternatives. More specifically, if

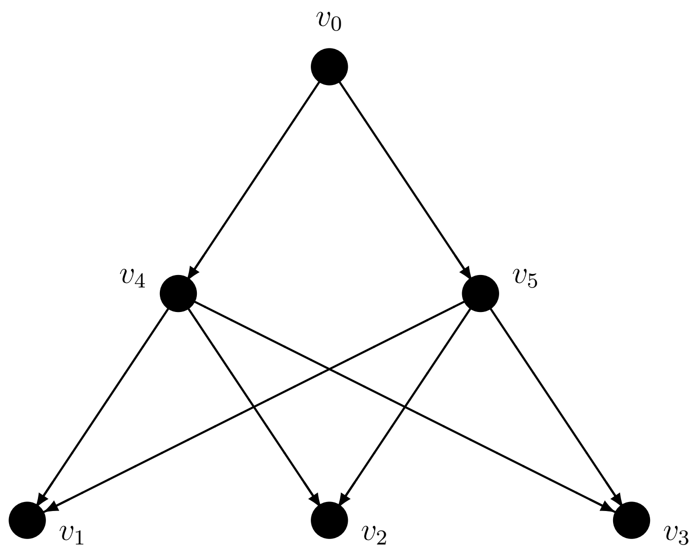

Figure 1: A simple network

we consider a path $\mathcal{P} = (v_{i_0}, v_{i_1}), (v_{i_1}, v_{i_2}), \ldots, (v_{i_{P-1}}, v_{i_P})$ connecting $v_0 = v_{i_0}$ and $v_i = v_{i_P}$, the parameter $\omega_{\mathcal{P}}$ associated with the path is

$$\omega_{\mathcal{P}} = \prod_{j=1}^{P} (\alpha_{(v_{i_{j-1}}, v_{i_j})})^{\frac{\mu_i}{\mu_{i_{j-1}}}}. \tag{16}$$

By assumption, $\omega_{\mathcal{P}}$ must be non zero for at least one path between v_0 and v_i. If the Network GEV model could be written in terms of parameters ω, it is recommended to keep the parameters α in the formulation. In that case, the intuitive interpretation of the model based on the network structure is maintained. If $\alpha_{(i,j)}$ is interpreted as the level at which node v_j "belongs to" node v_i, like for the Cross Nested logit model, a natural normalization would be to impose

$$\sum_{i \in \text{pred}(j)} \alpha_{(i,j)} = 1 \quad \forall j. \tag{17}$$

In particular, for any node j with only one predecessor i, we have $\alpha_{(i,j)} = 1$.

The estimation of a Network GEV model can be performed using any package able to estimate a GEV model. The freeware BIOGEME (Bierlaire, 2001b) is currently being extended to Network GEV models.

5 Conclusion and Future research

In this paper, we have proposed a new formulation of GEV models based on a network representation. It allows to design a GEV model based on intuitive interpretation of the application, similarly to the development of trees for Nested Logit models. We have proven in Section 3 that any Network GEV model is indeed a GEV model and, consequently, complies with random utility theory. The Multinomial logit, the Nested logit and the Cross Nested logit models are all Network GEV models. The issue of estimability has been discussed, but requires a full theoretical and empirical analysis which is out of the scope of this paper. An interesting question that should be investigated is the equivalence between GEV and Network GEV. Namely, can any GEV model be represented as a Network GEV model?

6 Acknowledgments

The author likes to thank Jean-Albert Ferrez and Alain Prodon for their help in proving the technical lemmas.

A Technical lemmas

Lemma 2 *We denote by $\mathcal{P}_k$ the set of partitions of the indices set $\{1,\dots,k\}$. Given a partition P belonging to $\mathcal{P}_k$, composed of p sets of indices, we define*

$$S^p(x) = \prod_{i=0}^{p-1}(\beta - i)G(x)^{\beta-p}, \tag{18}$$

Given a set R containing r indices, we define

$$D_R = \frac{\partial^r G(x)}{(\partial x_i)_{i\in R}}. \tag{19}$$

Then, we have

$$\frac{\partial^k}{\partial x_1 \dots \partial x_k} G(x)^\beta = \sum_{P\in\mathcal{P}_k} S^p \prod_{R\in P} D_R. \tag{20}$$

Proof. The proof is by induction. The cases $k = 1, 2, 3$ are obvious (see (33), (34), (35)). Therefore, we assume that the result is true for k, and we prove it is true also for $k+1$. We first compute

$$\frac{\partial^{k+1}}{\partial x_1 \dots \partial x_{k+1}} G(x)^\beta = \frac{\partial}{\partial x_{k+1}} \left(\frac{\partial^k}{\partial x_1 \dots \partial x_k} G(x)^\beta \right) \tag{21}$$

that is

$$\begin{aligned} &\frac{\partial^{k+1}}{\partial x_1 \dots \partial x_{k+1}} G(x)^\beta \\ &= \frac{\partial}{\partial x_{k+1}} \left(\sum_{P\in\mathcal{P}_k} S^p \prod_{R\in P} D_R \right) \qquad (22) \\ &= \sum_{P\in\mathcal{P}_k} \left[\left(\frac{\partial}{\partial x_{k+1}} S^p \right) \prod_{R\in P} D_R + S^p \frac{\partial}{\partial x_{k+1}} \prod_{R\in P} D_R \right], \end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial x_{k+1}} S^p &= \prod_{i=0}^{p-1}(\beta - i)(\beta - p)G(x)^{\beta-(p+1)}\frac{\partial G}{\partial x_{k+1}} \\ &= S^{p+1}\frac{\partial G}{\partial x_{k+1}}\end{aligned} \tag{23}$$

and

$$\frac{\partial}{\partial x_{k+1}} \prod_{R\in P} D_R = \sum_{R\in P} \frac{\partial}{\partial x_{k+1}} D_R \prod_{T\in P, T\neq R} D_T. \tag{24}$$

Consequently,

$$\begin{gathered}\frac{\partial^{k+1}}{\partial x_1 \ldots \partial x_{k+1}} G(x)^\beta = \\ \sum_{P\in\mathcal{P}_k}\left[\left(S^{p+1}\frac{\partial G}{\partial x_{k+1}}\right)\prod_{R\in P} D_R + S^p \sum_{R\in P} D_{R\cup\{k+1\}} \prod_{T\in P, T\neq R} D_T\right].\end{gathered} \tag{25}$$

Then, we prove the result by obtaining (25) directly from (20). Using (38) in (20), we have

$$\begin{aligned}&\frac{\partial^{k+1}}{\partial x_1 \ldots \partial x_{k+1}} G(x)^\beta = \\ &\sum_{i=1}^{p}\left[S^{p+1}\prod_{R\in P\cup\{k+1\}} D_R + \sum_{\ell=1}^{n_i} S^p \prod_{R\in P_{i,\ell}^{k+1}} D_R\right] \\ &\sum_{i=1}^{p}\left[S^{p+1}\frac{\partial G}{\partial x_{k+1}}\prod_{R\in P} D_R + S^p\sum_{\ell=1}^{n_i} \prod_{R\in P_{i,\ell}^{k+1}} D_R\right].\end{aligned} \tag{26}$$

We finally use the definition (39) of $P_{i,\ell}^{k+1}$ to obtain (25) and prove the result. $\square$

Lemma 3 *If $0 < \beta \leq 1$ and $G(x)$ is non-negative, the sign of $S^p(x)$, defined by (18), is non-positive if p is even and non-negative if p is odd, that is*

$$(-1)^p S^p(x) \leq 0. \tag{27}$$

Proof. From (18), we have

$$S^p(x) = \beta \prod_{i=1}^{p-1} (\beta - i) G(x)^{\beta - p}. \tag{28}$$

As β and $G(x)$ are non-negative, the sign of $S^p(x)$ is the sign of $\prod_{i=1}^{p-1}(\beta - i)$. If $\beta = 1$, then $S^p(x) = 0$. If $\beta < 1$, then all factors of $\prod_{i=1}^{p-1}(\beta - i)$ are negative. In both cases, (27) is trivially verified. □

Lemma 4 *Let P be a partition of the indices set $\{1, \ldots, k\}$, composed of p subsets. Then, the sign of $\prod_{R \in P} D_R$, where D_R is defined by (19) depends on the parity of k and p, as reported in Table 2*

	k even	k odd
p even	≥ 0	≤ 0
p odd	≤ 0	≥ 0

Table 2: Sign of $\prod_{R \in P} D_R$

Proof. As G is a GEV function, we have that $(-1)^r D_R \leq 0$ or, equivalently,

$$(-1)^{r+1} D_R \geq 0, \tag{29}$$

where r is the number of indices in R. Therefore, the sign of $\prod_{R \in P} D_R$ is also the sign of

$$\prod_{R \in P} (-1)^{r+1}. \tag{30}$$

We prove the result by induction. If P contains only one set, that is $p = 1$ and $R = \{1, \ldots, k\}$, then (30) becomes $(-1)^{k+1}$. Noting that $p = 1$ is odd, the result is trivially verified in that case. Now, we assume that the result is true if P is composed of p sets, and prove it for $p + 1$. If P is composed of $p + 1$ sets, we select one set R^* in P. In that case, (30) is decomposed as

$$\prod_{R \in P} (-1)^{r+1} = (-1)^{r^*+1} \prod_{R \in P, R \neq R^*} (-1)^{r+1}, \tag{31}$$

where $\prod_{R \in P, R \neq R^*} (-1)^{r+1}$ corresponds to the case where we have a partition of a set of $k - r^*$ indices into p sets. All possible cases are considered in Table 3. Column (1) contains the parity of $k - r^*$. It is the same as the parity of k if and only if r^* is even. Column (2) contains the parity of p which is directly deduced from the parity of $p + 1$. Column (3) contains the sign of $(-1)^{r^*+1}$, directly deduced from the parity of r^*. Column (4) is obtained by the recursion assumption, using the parity of $k - r^*$ and p in Table (2). Finally, column (5) is obtained by multiplying columns (3) and (4). The signs in column (5) are the same as the signs in Table 2, and the result is proven.

$\square$

B Derivatives

We provide here the analytical value for

$$\frac{\partial^k}{\partial x_1 \ldots \partial x_k} G(x)^\beta \tag{32}$$

k	$p+1$	r^*	(1) $k-r^*$	(2) p	(3) $(-1)^{r^*+1}$	(4) $\prod_{R\in P, R\neq R^*}(-1)^{r+1}$	(5) $\prod_{R\in P}(-1)^{r+1}$
even	even	even	even	odd	≤ 0	≤ 0	≥ 0
even	even	odd	odd	odd	≥ 0	≥ 0	≥ 0
even	odd	even	even	even	≤ 0	≥ 0	≤ 0
even	odd	odd	odd	even	≥ 0	≤ 0	≤ 0
odd	even	even	odd	odd	≤ 0	≥ 0	≤ 0
odd	even	odd	even	odd	≥ 0	≤ 0	≤ 0
odd	odd	even	odd	even	≤ 0	≤ 0	≥ 0
odd	odd	odd	even	even	≥ 0	≥ 0	≥ 0

Table 3: Sign of (31)

for $k = 1, 2, 3$. For $k = 1$, we have

$$\frac{\partial}{\partial x_1} G(x)^\beta = \beta G(x)^{\beta-1} \frac{\partial G}{\partial x_1} \tag{33}$$

For $k = 2$, we have

$$\frac{\partial^2}{\partial x_1 \partial x_2} = \beta(\beta-1)G(x)^{\beta-2} \frac{\partial G}{\partial x_1}\frac{\partial G}{\partial x_2} + \beta G(x)^{\beta-1} \frac{\partial^2 G}{\partial x_1 \partial x_2}. \tag{34}$$

For $k = 3$, we have

$$\begin{aligned} \tfrac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} &= \beta(\beta-1)(\beta-2)G(x)^{\beta-3} \tfrac{\partial G}{\partial x_1}\tfrac{\partial G}{\partial x_2}\tfrac{\partial G}{\partial x_3} \\ &+ \beta(\beta-1)G(x)^{\beta-2}\left(\tfrac{\partial G}{\partial x_1}\tfrac{\partial^2 G}{\partial x_2 \partial x_3} + \tfrac{\partial G}{\partial x_2}\tfrac{\partial^2 G}{\partial x_1 \partial x_3} + \tfrac{\partial G}{\partial x_3}\tfrac{\partial^2 G}{\partial x_1 \partial x_2}\right) \\ &+ \beta G(x)^{\beta-1} \tfrac{\partial^3 G}{\partial x_1 \partial x_2 \partial x_3} \end{aligned} \tag{35}$$

C Set partitions

The set $\mathcal{P}_k$ of partitions of an index set $\{1,\ldots,k\}$ is constructed recursively. By definition, we impose that $\mathcal{P}_0 = \emptyset$. We have also that $\mathcal{P}_1 = \{\{\{1\}\}\}$. In general, assume that

$$\mathcal{P}_k = \bigcup_{i=1}^{p} P_i^k \tag{36}$$

where

$$P_i^k = \bigcup_{j=1}^{n_i} R_j^i \tag{37}$$

and R_j is a subset of $\{1,\ldots,k\}$. For each partition P_i^k in $\mathcal{P}_k$, we build n_i+1 partitions of $\mathcal{P}_{k+1}$. The first is obtained simply by adding the singleton $\{k+1\}$ to P_i^k. All the other partitions are obtained by replacing each index set R_j^i, one at a time, by $R_j^i \cup (k+1)$. Consequently, we have

$$\mathcal{P}_{k+1} = \bigcup_{i=1}^{p} \left[\left(P_i^k \cup \{k+1\}\right) \bigcup_{\ell=1}^{n_i} P_{i,\ell}^{k+1} \right], \tag{38}$$

where

$$P_{i,\ell}^{k+1} = \{R_\ell^i \cup (k+1)\} \cup \bigcup_{\substack{j=1 \\ j\neq\ell}}^{n_i} R_j^i. \tag{39}$$

For example, as $\mathcal{P}_1$ contains one partition $P_1^1 = \{R_1^1\}$, where $R_1^1 = \{1\}$, we have

$$\begin{aligned} \mathcal{P}_2 &= \textstyle\bigcup_{i=1}^{1} \left[(P_i^1 \cup \{2\}) \bigcup_{\ell=1}^{1} P_{i,\ell}^2 \right], \\ &= P_1^1 \cup \{2\} \cup P_{1,1}^2 \\ &= \{\{1\}\{2\}\} \cup P_{1,1}^2 \end{aligned} \tag{40}$$

where

$$P_{1,1}^2 = \{R_1^1 \cup 2\} \cup \bigcup_{\substack{j=1 \\ j\neq 1}}^{1} R_j^1 = \{\{1,2\}\}. \tag{41}$$

Therefore, $\mathcal{P}_2 = \{\{\{1\}\{2\}\}, \{\{1,2\}\}\}$. Denoting $P_1^2 = \{R_1^1, R_2^1\}$, with $R_1^1 = \{1\}$ and $R_2^1 = \{2\}$ and $P_2^2 = \{R_1^2\}$,where $R_1^2 = \{1,2\}$, we can compute $\mathcal{P}_3$.

$$\begin{aligned}
\mathcal{P}_3 &= \bigcup_{i=1}^{2} \left[(P_i^2 \cup \{3\}) \bigcup_{\ell=1}^{n_i} P_{i,\ell}^3\right], \\
&= \left[(P_1^2 \cup \{3\}) \bigcup_{\ell=1}^{2} P_{1,\ell}^3\right] \bigcup \left[(P_2^2 \cup \{3\}) \bigcup_{\ell=1}^{1} P_{2,\ell}^3\right] \\
&= \left[(P_1^2 \cup \{3\}) \cup P_{1,1}^3 \cup P_{1,2}^3\right] \bigcup \left[(P_2^2 \cup \{3\}) \cup P_{2,1}^3\right]
\end{aligned} \tag{42}$$

where

$$\begin{aligned}
P_{1,1}^3 &= \{R_1^1 \cup 3\} \cup \bigcup_{\substack{j=1 \\ j\neq 1}}^{2} R_j^1 \\
&= \{R_1^1 \cup 3\} \cup R_2^1 \\
&= \{\{1\} \cup 3\} \cup \{2\}, \\
&= \{\{1,3\}\{2\}\} \\
P_{1,2}^3 &= \{\{2,3\}\{1\}\} \\
P_{2,1}^3 &= \{\{1,2,3\}\}
\end{aligned} \tag{43}$$

Consequently,

$$\begin{aligned}
&\mathcal{P}_3 = \{ \\
&\{\{1\}\{2\}\{3\}\}, \\
&\{\{1,3\}\{2\}\}, \\
&\{\{2,3\}\{1\}\}, \\
&\{\{1,2\}\{3\}\}, \\
&\{\{1,2,3\}\} \\
&\}
\end{aligned} \tag{44}$$

References

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions, *in* R. Hall (ed.), *Handbook of Transportation Science*, Kluwer, pp. 5–34.

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

Bierlaire, M. (1995). A robust algorithm for the simultaneous estimation of hierarchical logit models, *GRT Report 95/3*, Department of Mathematics, FUNDP.

Bierlaire, M. (2001a). A general formulation of the cross-nested logit model, *Proceedings of the 1st Swiss Transportation Research Conference*, Ascona, Switzerland.
URL: *http://www.strc.ch*

Bierlaire, M. (2001b). An introductory tutorial to BIOGEME.
URL: *http://rosowww.epfl.ch/mbi/biogeme*

Bierlaire, M. and Vandevyvere, Y. (1995). *HieLoW: the interactive user's guide*, Transportation Research Group - FUNDP, Namur.

Daly, A. (1987). Estimating "tree" logit models, *Transportation Research B* **21**(4): 251–268.

Daly, A. (2001). ????? Paper presented at the *Invitational Choice Symposium*, Asilomar, Ca.

McFadden, D. (1978). Modelling the choice of residential location, *in* A. Karlquist *et al.* (ed.), *Spatial interaction theory and residential location*, North-Holland, Amsterdam, pp. 75–96.

Papola, A. (2000). Some development of the cross-nested logit model, *Proceedings of the 9th IATBR Conference*.

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

Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models, *TRB* **35**(7): 643–666.

Vovsha, P. (1997). Cross-nested logit model: an application to mode choice in the Tel-Aviv metropolitan area, Transportation Research Board, 76th Annual Meeting, Washington DC. Paper #970387.

Vovsha, P. and Bekhor, S. (1998). The link-nested logit model of route choice: overcoming the route overlapping problem, *Transportation Research Record* **1645**: 133–142.

Wen, C.-H. and Koppelman, F. S. (2001). The generalized nested logit model, *Transportation Research B* **35**(7): 627–641.