A general and operational representation of GEV models

Andrew Daly* Michel Bierlaire†

Report RO-030502
May 2, 2003

Abstract

Generalised Extreme Value models provide an interesting theoretical framework to develop closed-form random utility models. Unfortunately, few instances of this family are to be found as operational models in practice. The form of the model, based on a generating function G which must satisfy specific properties, is rather complicated. Fundamentally, it is not an easy task to translate an intuitive perception of the correlation structure by the modeller into a concrete G function. And even if the modeller succeeds in proposing a new G function, the task of proving that it indeed satisfies the properties is cumbersome.

The main objectives of this paper are (i) to provide a general theoretical foundation, so that the development of new GEV models will be easier in the future, and (ii) to propose an easy way of generating new GEV models without a need for complicated proofs. Our technique requires only a network structure capturing the underlying correlation of the choice situation under consideration. If the network complies with some simple conditions, we show how to build an associated

*RAND Europe and Institute for Transport Studies, University of Leeds, UK. Email: daly@rand.org

†Ecole Polytechnique Fédérale de Lausanne, Dpt. of Mathematics, CH-1015 Lausanne, Switzerland. Email: michel.bierlaire@epfl.ch

model. We prove that it 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 specific instances of our class of models. So are the recent GenL model, combining choice set generation and choice model and some specialised compound models used in recent work. Probability, expected maximum utility and elasticity formulae for the class of models are provided.

Preferred citation:
Daly, A. and Bierlaire, M. (2003). A general and operational representation of GEV models, *Technical Report RO-030502*, Institute of Mathematics, Operations Research Group ROSO, EPFL, Lausanne, Switzerland.

1 Introduction

Discrete choice models play a major role in many fields involving a human dimension, including econometrics, marketing research and transportation demand analysis. Their attractive and firmly based theoretical properties, and their flexibility to capture a wide range of situations, provide a vast area of interest for both researchers and practitioners, of which only a limited part has yet been exploited. The theory on Generalised Extreme Value (GEV) models was introduced by McFadden (1978) (see also McFadden, 1981). It provides 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, 2001, Wen and Koppelman, 2001, Swait, 2001) has slightly extended the number of GEV models used in practice. Specialised compound GEV models (Bhat, 1998, Whelan et al., 2002) have further extended the GEV forms used.

Dagsvik (1994) indicates that there is a huge variety of GEV models, fitting effectively every possible random utility model 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 hybrid choice models, which are in many cases easier to construct (McFadden, 2000, and Ben-Akiva et al., 2002 are examples of a burgeoning literature). However, GEV models still offer considerable advantages. Indeed, their closed form formulation simplifies the computation of the model and its derivatives both for estimation and applications. The GEV function itself, a summary of the total utility enjoyed by consumers, is a convenient and theoretically sound measure which can be applied in further modelling or evaluation. Moreover, several of the GEV models can be proved to converge in tranport planning applications with conventional assignment procedures (Prashker and Bekhor, 2000). Finally, the hybrid models, currently based on the combination of logit models and multinomial probit models, can be easily extended to hybrid GEV models, combining a GEV model with a probit model or other mixing distribution.

The complexity of the assumptions associated with the GEV models (see Section 2) is probably the cause of the few concrete instances from the GEV model family proposed in the literature.

In this paper, we propose a general and operational representation, providing an intuitive way of generating a wide class of concrete GEV models. This representation is based on the Recursive Nested Extreme Value (RNEV) model proposed by Daly (2001b), and the network GEV model proposed by Bierlaire (2002). The RNEV was originally designed as a generalisation of the Cross-Nested logit model, in which multiple layers of nests are allowed. The Network GEV model generalises the use of trees to represent Nested Logit models (Ben-Akiva and Lerman, 1985, Daly, 1987) to a network representation. The advantages of our approach are the following.

- The GEV inheritance results allow the generation of new GEV models from existing ones.
- The GEV Network representation allows complex correlation structures of actual modelling situations to be captured intuitively. 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.

- Properties established for the network model automatically apply to its many special cases.

The main objective of the paper is to provide a general theoretical foundation, making the development of new GEV models easier in the future. Indeed, in addition to the intuitive approach based on the network structure, any instance of that class of models is proven to be a GEV model and therefore, no further theoretical justification is required for such models.

After a short presentation of GEV models in Section 2, the GEV inheritance results are presented in Section 3. The modelling framework based on a network structure is formally defined in Section 4, where we prove that it belongs to the GEV family under weak conditions. Demand responses and elasticities are derived in Section 5. In Section 6, we describe some concrete instances, while estimation issues are discussed in Section 7. Finally, Appendix A provides a formal proof of the main GEV inheritance theorem, and Appendix B provides the derivatives of the G function as required by optimisation packages for model estimation. Appendix C provides some definitions, notations and examples about set partioning useful in the proofs. Finally, Appendix D provides the derivation of some classical results related to GEV models.

2 The GEV model

The Generalised Extreme Value (GEV) model has been derived from the random utility paradigm 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 with alternative i, and G is a μ-GEV function. A μ-GEV function is a differentiable function defined on $\mathbb{R}_+^J$ with the following properties:

1. $G(y) \geq 0$ for all $y \in \mathbb{R}_+^J$,

2. G is homogeneous of degree $\mu > 0$, that is $G(\lambda y) = \lambda^\mu G(y)$, for $\lambda > 0$,

3. $\lim_{y_i \to +\infty} G(y_1, \ldots, y_i, \ldots, y_J) = +\infty$, for each $i = 1, \ldots, J$,

4. The mixed partial derivatives of G exist and are continuous. Moreover, 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 $i_1, \ldots, i_k \in \mathcal{J} = \{1, \ldots, J\}$, we have

$$(-1)^k D_{\mathcal{K}}(x) \leq 0, \ \forall x \in \mathbb{R}_+^J, \tag{2}$$

where $\mathcal{K} = \{i_1, \ldots, i_k\} \subseteq \mathcal{J}$

$$D_{\mathcal{K}}(x) = \frac{\partial^k G}{\partial x_{i_1} \ldots \partial x_{i_k}}(x). \tag{3}$$

For future reference, we say that a function satisfying this property is *GEV-differentiable.*

Although this condition is never stated in the literature, it is also required that $G(x) \neq 0$.

Euler's formula and the homogeneity of G can be invoked to transform (1) into a form similar to the multinomial logit model (see McFadden, 1978 and Ben-Akiva and Lerman, 1985):

$$P(i|C) = \frac{e^{V_i + \log G_i(\ldots)}}{\sum_{j=1}^J e^{V_j + \log G_j(\ldots)}}, \tag{4}$$

where $G_i = \partial G / \partial y_i$. Also, the expected maximum utility corresponding to choice in the model generated by G is

$$\bar{U} = \frac{\log G(x) + \gamma}{\mu} \tag{5}$$

where γ is Euler's constant, and

$$P(i|C) = \frac{\partial \bar{U}}{\partial V_i}. \tag{6}$$

Note that, in general, GEV properties 1, 2 and 3 are easily verified in practice. GEV-differentiability can be cumbersome to verify, so that many authors avoid doing so explicitly. Statement like "it is easy to prove that" are regularly found in the literature. In the case of the Cross-Nested Logit model, for example, the proof is definitely not trivial (see Bierlaire, 2001).

3 GEV inheritance

In this section, we provide some theoretical results enabling the derivation of GEV functions from other GEV functions. Such results are designed to facilitate the derivation of new GEV functions and, therefore, new GEV models. Theorem 1 deals with the GEV-inheritance by linear combinations of GEV functions. Theorem 4 deals with the GEV-inheritance when a GEV function is raised to a power. The combination of these two results yields Theorem 7, the main GEV Inheritance Theorem. Corollaries 2, 5 and 8 deal with the expected maximum utility for each case respectively, while Corollaries 3, 6 and 9 give the probability models.

Theorem 1 *Let* $\mathbb{R}^{J_i}$ *be* p *subspaces spanning* $\mathbb{R}^J$. *For any vector* $x \in \mathbb{R}^J$, $[x]_i$ *denotes the projection of* x *on* $\mathbb{R}^{J_i}$. *Let* $G^i : \mathbb{R}_+^{J_i} \longrightarrow \mathbb{R}$, $i = 1, \ldots, p$ *be* μ*-GEV functions. Then, the function*

$$G : \mathbb{R}_+^J \longrightarrow \mathbb{R} : x \rightsquigarrow G(x) = \sum_{i=1}^{p} \alpha_i G^i([x]_i) \tag{7}$$

is also a μ*-GEV function if* $\alpha_i > 0$, $i = 1, \ldots, p$.

Proof. All four properties are obviously verified, as $\alpha_i > 0$. □

Corollary 2 *Under the hypotheses of Theorem 1, the expected maximum utility corresponding to* G *is*

$$\bar{U} = \frac{\log \sum_{i=1}^{p} \alpha_i e^{\mu \bar{U}_i}}{\mu}, \tag{8}$$

where $\bar{U}_i$ *is the expected maximum utility corresponding to each* G^i*.*

Proof. From (5), we have

$$G^i([x]_i) = e^{\mu \bar{U}_i - \gamma}$$

so that

$$G(x) = \sum_{i=1}^{p} \alpha_i e^{\mu \bar{U}_i - \gamma} = e^{-\gamma} \sum_{i=1}^{p} \alpha_i e^{\mu \bar{U}_i}. \tag{9}$$

From (5), we obtain (8). □

Corollary 3 *Under the hypotheses of Theorem 1, we have*

$$P(k) = \sum_{i=1}^{p} \frac{\alpha_i e^{\mu \bar{U}_i}}{\sum_j \alpha_j e^{\mu \bar{U}_j}} P_i(k) \tag{10}$$

where $P(k)$ *is the probability of choosing alternative* k *based on the model generated by* G *and* $P_i(k)$ *the probability based on the model generated by* G_i*.*

Proof. From (6), we have

$$P(k) = \frac{\partial \bar{U}}{\partial V_k} = \sum_{i=1}^{p} \frac{\partial \bar{U}}{\partial \bar{U}_i} \frac{\partial \bar{U}_i}{\partial V_k} = \sum_{i=1}^{p} \frac{\partial \bar{U}}{\partial \bar{U}_i} P_i(k). \tag{11}$$

We use Corollary 2 to complete the proof by differentiating (8).
□

Theorem 4 *Let* $G : \mathbb{R}_+^J \longrightarrow \mathbb{R}$ *be a* μ*-GEV function. Then* G^β *is a* $(\mu\beta)$*-GEV function if* $0 < \beta \leq 1$*.*

Proof.

1. G^β is obviously non negative.
2. From the μ-homogeneity of G, we have $G(\lambda y)^\beta = (\lambda^\mu G(y))^\beta = \lambda^{\mu\beta} G(y)^\beta$ and, therefore, G^β is homogeneous of degree $\mu\beta$.
3. The limit property holds as a consequence of $\beta > 0$.

4. We provide here a short indication of the validity of the Theorem. A detailed proof is given in Appendix A. Using basic calculus and induction on the order of differentiation, it is easy to show that $D^*_{\mathcal{J}}(x)$, defined similarly to (3) for $G^* = G^\beta$, is composed of terms of the form

$$kG(x)^q \prod_{\mathcal{K}} G_{\mathcal{K}}(x),$$

where k is a numerical constant, $q \leq 0$, and $\mathcal{K} \subseteq \mathcal{J}$. For the first derivative, we have

$$\begin{aligned} D^*_{\{1\}}(x) &= \tfrac{\partial}{\partial x_1} G(x)^\beta \\ &= \beta G(x)^{\beta-1} \tfrac{\partial G}{\partial x_1}(x) \\ &= \beta G(x)^{\beta-1} G_{\{1\}}(x) \end{aligned} \tag{12}$$

This is non-negative because β and G are positive by assumption and $G_{\{1\}}$ is non-negative because G is a GEV function. Note also that $\beta - 1 \leq 0$, so that the term is of the form specified. Moreover, when a term of the specified form is differentiated, we obtain a further series of terms of the required form.

It is now sufficient to show that differentiation of each of these terms leads to a change of sign. The differentiation of the component in $G(x)^q$ yields a term

$$kqG(x)^{q-1} \prod_{\mathcal{K}} G_{\mathcal{K}}(x),$$

which is zero or of opposite sign to the original because G is positive while q is non-positive. Differentiation of the terms within the product leads to a change in sign because one of the terms in the product is increased by one in its order of differentiation, causing a change in sign because of the GEV-differentiability of G.

The fact that the first derivative has the correct sign and subsequent derivatives alternate in sign completes the proof.

□

Corollary 5 *Under the hypotheses of Theorem 4, we have*

$$\bar{U}^* = \bar{U} + \frac{\gamma}{\mu}(\frac{1}{\beta} - 1) \tag{13}$$

where $\bar{U}^*$ *is the expected maximum utility for the model generated by* G^β*, and* $\bar{U}$ *for the model generated by* G*.*

Proof. Let $\bar{U}^*$ be the exected maximum utility for the model based on G^β. We have

$$\bar{U}^* = \frac{\log G^\beta + \gamma}{\mu\beta} = \frac{\beta \log G + \gamma}{\mu\beta} = \frac{\log G}{\mu} + \frac{\gamma}{\mu\beta} + \frac{\gamma}{\mu} - \frac{\gamma}{\mu}, \tag{14}$$

and the result follows directly. □

Corollary 6 *Under the hypotheses of Theorem 4, the probability of choosing an alternative given by the model based on* G^β *is the same as the probability given by the model based on* G*.*

Proof. Denoting by $P^*(k)$ the probability given by the model based on G^β, we use (6) to obtain

$$P^*(k) = \frac{\partial \bar{U}^*}{\partial V_k} = \frac{\partial \bar{U}^*}{\partial \bar{U}} \frac{\partial \bar{U}}{\partial V_k} = P(k), \tag{15}$$

where $P(k)$ is the probability given by the model based on G. The result is obtained by differentiating (13). □

Theorems 1 and 4 can be combined into a more general result, formalised in the next theorem.

Theorem 7 ***GEV Inheritance*** *Let* $\mathbb{R}^{J_i}$ *be* p *subspaces spanning* $\mathbb{R}^J$*. For any vector* $x \in \mathbb{R}^J$*, denote by* $[x]_i$ *the projection of* x *on* $\mathbb{R}^{J_i}$*. Let* $G^i : \mathbb{R}_+^{J_i} \longrightarrow \mathbb{R}$*,* $i = 1, \ldots, p$ *be* μ_i*-GEV functions. Then, the function*

$$G : \mathbb{R}_+^J \longrightarrow \mathbb{R} : x \rightsquigarrow G(x) = \sum_{i=1}^{p} \alpha_i G^i([x]_i)^{\frac{\mu}{\mu_i}} \tag{16}$$

is a μ*-GEV function if* $\alpha_i > 0$ *and* $0 < \mu \leq \mu_i$*,* $i = 1, \ldots, p$*.*

Proof. It is a direct consequence of Theorems 1 and 4. □

Corollary 8 *Under the hypotheses of Theorem 7, we have*

$$\bar{U} = \frac{\gamma + \log \sum_{i=1}^{p} \alpha_i e^{\mu\left(\bar{U}_i - \frac{\gamma}{\mu_i}\right)}}{\mu}, \tag{17}$$

where $\bar{U}$ *is the expected maximum utility for the model generated by* G*, and* $\bar{U}_i$ *for the model generated by* G^i*.*

Proof. Using Corollary 5, the expected maximum utility for the model based on $(G^i)^{\mu/\mu_i}$ is

$$\bar{U}_i + \frac{\gamma}{\mu_i}(\frac{\mu_i}{\mu} - 1) = \bar{U}_i + \frac{\gamma}{\mu} - \frac{\gamma}{\mu_i}. \tag{18}$$

The result is a direct consequence of Corollary 2. □

Corollary 9 *Under the hypotheses of Theorem 7, we have*

$$P(k) = \sum_{i=1}^{p} \Omega_i P_i(k) \tag{19}$$

where

$$\Omega_i = \frac{\alpha_i e^{\mu(\bar{U}_i - \gamma/\mu_i)}}{\sum_j \alpha_j e^{\mu(\bar{U}_j - \gamma/\mu_j)}}, \tag{20}$$

$P(k)$ *is the probability of choosing alternative* k *based on the model generated by* G *and* $P_i(k)$ *the probability based on the model generated by* G_i*.*

Proof. It is a direct consequence of Corollaries 3, 5 and 6. □

Theorem 7 is an extension of a property mentioned by McFadden (2000): *If* $H^A(w_A)$ *and* $H^B(w_B)$ *are GEV generating functions in* w_A *and* w_B*, respectively, and if* $s \geq 1$*, then* $H^C(w_C) = H^A(w_A)^{1/s} + H^B(w_B)$ *is a GEV generating function in* w_C*.* Daly (2001b) developed the Recursive Nested Extreme Value (RNEV) model based on recursive application of (16), also giving a different but less elegant proof of Theorem 7.

4 GEV network

We now exploit the theoretical results presented above to propose an operational representation of the RNEV model (Daly, 2001b) based on the network structure proposed by Bierlaire (2002).

Let (N, E) be a finite non-empty directed graph, where N is the set of vertices, or nodes, and E the set of arcs (N and E are finite). Each arc (i, j) is associated with a non-negative parameter $\alpha_{ij} > 0$, so that the directed graph is a network. We assume that the network does not contain any circuit.

We consider two special subsets of nodes:

1. $\mathcal{R}$ is the set of nodes with no predecessor,

2. $\mathcal{C}$ is the set of J nodes with no successor.

Because the network is non-empty and circuit-free, both $\mathcal{R}$ and $\mathcal{C}$ are non-empty.

Our objective is to define a family of GEV models based on such a network, where each node in $\mathcal{C}$ is associated with an alternative and therefore, $\mathcal{C}$ represents the choice set. All other nodes represent nests. Nodes in $\mathcal{R}$ are called *roots*, and can be either alternatives or nests. For most practical purposes, most models will be defined by a network with a single root. Actually, such a network can be trivially derived from a multi-root network by adding a node and connecting it to all nodes in $\mathcal{R}$. The additional node is the only root of the new network, and is a nest.

We associate with each node v_i a parameter $\mu_i > 0$, and we recursively define a subspace $\mathbb{R}^{J_i}$ of $\mathbb{R}^J$, and a function $G^i : \mathbb{R}_+^{J_i} \longrightarrow \mathbb{R}$. If $v_i \in \mathcal{C}$, then the subspace is $\mathbb{R}$ and

$$G^i : \mathbb{R} \longrightarrow \mathbb{R} : G^i(x_i) = x_i^{\mu_i}, \tag{21}$$

with $\mu_i > 0$. If $v_i \notin \mathcal{C}$, then the subspace is

$$\mathbb{R}^{J_i} = \text{span}_{v_j \in S(v_i)} < \mathbb{R}^{J_j} >, \tag{22}$$

where $S(\nu_i)$ is the set of successors of ν_i, and

$$G^i : \mathbb{R}^{J_i} \longrightarrow \mathbb{R} : G^i(x) = \sum_{\nu_j \in S(\nu_i)} \alpha_{ij} G^j(x)^{\frac{\mu_i}{\mu_j}}. \tag{23}$$

Theorem 10 *The function G^i associated with any node ν_i of a GEV network is a μ_i-GEV function if, for all $\nu_j \notin \mathcal{C}$,*

$$\mu_j \leq \mu_k \quad \forall k \textit{ such that } \nu_k \in S(\nu_j). \tag{24}$$

Proof. The proof is by induction. At each step of the induction, we consider a set of nodes $\hat{N}_k$ and show that the theorem is verified for all nodes in that set. At the last step, we show that all nodes of the network are in $\hat{N}_k$.

For $k = 0$, we define $\hat{N}_0 = \mathcal{C}$. The G^i functions defined by (21) are clearly GEV-functions. Let $\hat{N}_k$ be a set of nodes ν_i such that the associated function G^i is a μ_i-GEV function. If $\hat{N}_k = N$, the result is proven. If not, we build $\hat{N}_{k+1}$ by adding to $\hat{N}_k$ all immediate predecessors of its elements, not already in the set, that is

$$\hat{N}_{k+1} = \hat{N}_k \cup \left\{ \nu \in N \setminus \hat{N}_k \mid w \in S(\nu) \Longrightarrow w \in \hat{N}_k \right\}.$$

For all nodes that have been added, Theorem 7 applies to the definition of G^i in (23), as condition (24) is verified. Consequently, the GEV property for the added nodes is a direct consequence of the GEV property for their successors in $\hat{N}_k$.

The induction terminates when $\hat{N}_k = \hat{N}_{k+1}$, that is when no node can be added anymore. Consider the set $N^* = N \setminus \hat{N}_k$ of remaining nodes. Any node in N^* must have successors, because they are not members of $\mathcal{C}$, which is a subset of $\hat{N}_k$. Not all the successors can be in $\hat{N}_k$, because the induction has terminated. Consequently, all the nodes in N^* have some of their successors in N^* and therefore cycles exist in N^*, contrary to the specification of the network itself. Therefore, N^* must be empty, $\hat{N}_k = N$, and the theorem is proven.

□

Thus for any node of a GEV network, the GEV model of choice given by (1) applies, and this model is consistent with McFadden's GEV theory and hence with utility maximisation. This result has been proven by Bierlaire (2002) in a different way. The function $\bar{U}_i = (\log G_i + \gamma)/\mu_i$ at each node ν_i is a measure of the average utility at that node derived from the relevant alternatives. For $\nu_i \in \mathcal{C}$, we obviously have

$$\bar{U}_i = \frac{\log e^{\mu_i V_i} + \gamma}{\mu_i} = V_i + \frac{\gamma}{\mu_i}, \tag{25}$$

and

$$P_i(k) = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

It is important to emphasise that the conditions required on the network (finite, non-empty, circuit-free) are easily verified in practice. The role of the modeller is now to design a network structure that adequately represents the underlying correlation structure of the model. No more proof is necessary.

Also, the network representation can always be used to generalise any GEV model, as the GEV properties propagate across the network, according to the GEV inheritance theorems. It suffices to redefine the role of the nodes with no successor, and to assign a GEV model to each of them.

We now analyse the probabilities given by a single-root Network GEV model. We generalise (19) to obtain the recursive definition of the conditional probability model corresponding to a node j in the network

$$P_j(i) = \sum_{k \in S_j} \Omega_{jk} P_k(i), \tag{27}$$

where

$$\Omega_{jk} = \frac{\alpha_{jk} e^{\mu_j \bar{U}_k - \frac{\gamma}{\mu_k}}}{\sum_{\ell \in S_j} \alpha_{j\ell} e^{\mu_j \bar{U}_\ell - \frac{\gamma}{\mu_\ell}}} = \frac{\alpha_{jk} G^k}{\sum_{\ell \in S_j} \alpha_{j\ell} G^\ell}. \tag{28}$$

If there is one path $\mathcal{K}_{ij} = \{(k_1, k_2), (k_2, k_3), \ldots, (k_{K_{ij}}, k_{K_{ij}+1})\}$ connecting j to i that is, such that $k_1 = j$ and $k_{K_{ij}+1} = i$, then we can write

$$P_j(i) = \prod_{k=1}^{K_{ij}} P_k(k+1) = \prod_{k=1}^{K_{ij}} \frac{\partial \bar{U}_k}{\partial \bar{U}_{k+1}}. \tag{29}$$

If there are multiple paths, the probability is

$$P_j(i) = \sum_{\mathcal{K}_{ij}} \prod_{k=1}^{K_{ij}} P_k(k+1) = \sum_{\mathcal{K}_{ij}} \prod_{k=1}^{K_{ij}} \frac{\partial \bar{U}_k}{\partial \bar{U}_{k+1}}, \tag{30}$$

where $P_k(k+1)$ is the probability of choosing the composite alternative $k+1$ given by the model associated with node k, and is given by (27).

5 Demand responses and elasticities

The standard formula for the (point) elasticity of demand for alternative i with respect to an attribute x_j of alternative j can be stated as

$$\eta_{i,x_j} = \frac{\partial P(i)}{\partial x_j} \frac{x_j}{P(i)} = \frac{\partial P(i)}{\partial V_j} \frac{\partial V_j}{\partial x_j} \frac{x_j}{P(i)}. \tag{31}$$

When the utility function is linear in x, $\partial V_j / \partial x_j$ is constant and is equal to the coefficient of x_j, β_{x_j}.

The value $P(i)$ would usually be taken to be the population average, i.e. the market share for alternative i. For x_j, the population average is often taken, but it would also be possible to use the specific value facing each individual, i.e. to analyse the elasticity as a constant proportional change in the relevant attribute for each individual; the latter approach may not be appropriate if the value zero is possible.

Because the confusions relating to β_{x_j}, x_j and $P(i)$, it is much clearer to analyse model structures in terms of the responses of demand to changes in the utility of alternatives, i.e. $\partial P(i)/\partial V_j$. Eq. (31) can always be used to obtain elasticities.

We analyse demand responses within the context of the GEV inheritance theorems. Actually, it is relevant only for Theorems 1 and 7. Indeed, the inheritance addressed by Theorem 4 does not affect the probability and, consequently, the demand response. We first establish a simple technical lemma, easy to verify.

Lemma 11 *If we consider*

$$T_i^a = \alpha_i e^{\mu \bar{U}_i} \tag{32}$$

or

$$T_i^b = \alpha_i e^{\mu \bar{U}_i - \frac{\gamma}{\mu_i}}, \tag{33}$$

in both cases we have

$$\frac{\partial T_i^*}{\partial \bar{U}_m} = \begin{cases} \mu T_i^* & \textit{if } i = m, \\ 0 & \textit{otherwise}, \end{cases} \tag{34}$$

where the symbol $*$ *represents either* a *or* b*.*

Theorem 12 *Under the hypotheses of Theorem 1, the inheritance of demand responses is given by*

$$\frac{\partial P(k)}{\partial V_\ell} = \sum_{i=1}^{p} \left\{ \Omega_i \frac{\partial P_i(k)}{\partial V_\ell} + \mu P_i(k) \Omega_i \left(P_i(\ell) - \sum_{m=1}^{p} P_m(\ell) \Omega_m \right) \right\}, \tag{35}$$

where $\Omega_i = T_i^a / S$*,* $S = \sum_j T_j^a$*, and* T_i^a *is defined by (32).*

Proof. We write the probability (10) as

$$P(k) = \sum_i \Omega_i P_i(k), \tag{36}$$

We have that

$$\frac{\partial P(k)}{\partial V_\ell} = \sum_{i=1}^{p} \frac{\partial \Omega_i}{\partial V_\ell} P_i(k) + \Omega_i \frac{\partial P_i(k)}{\partial V_\ell}. \tag{37}$$

Then,

$$\frac{\partial \Omega_i}{\partial V_\ell} = \sum_{m=1}^{p} \frac{\partial \Omega_i}{\partial \bar{U}_m} \frac{\partial \bar{U}_m}{\partial V_\ell} = \sum_{m=1}^{p} \frac{\partial \Omega_i}{\partial \bar{U}_m} P_m(\ell).$$

As $\Omega_i = T_i^a / S$, we have

$$\frac{\partial \Omega_i}{\partial \bar{U}_m} = \frac{\partial T_i^a}{\partial \bar{U}_m} \frac{1}{S} - \frac{1}{S^2} \frac{\partial S}{\partial \bar{U}_m}.$$

Using Lemma 11 and simplifying the expression, we obtain that

$$\frac{\partial \Omega_i}{\partial \bar{U}_m} = \begin{cases} \mu \Omega_m (1 - \Omega_m) & \text{if } i = m \\ -\mu \Omega_i \Omega_m & \text{otherwise.} \end{cases}$$

Therefore,

$$\frac{\partial \Omega_i}{\partial V_\ell} = \mu \Omega_i \left(P_i(\ell) - \sum_{m=1}^{p} \Omega_m P_m(\ell) \right),$$

and (35) follows from (37). □

Corollary 13 *Under the hypotheses of Theorem 7, the inheritance of demand responses is given by (35) where* $\Omega_i = T_i^b / S$, $S = \sum_j T_j^b$, *and* T_i^b *is defined by (33).*

Proof. The proof is exactly the same as for Theorem 12, with T_b instead of T_a, because of Lemma 11. □

6 Specific instances

The framework combining the RNEV model and the GEV Network representation generalises many models proposed in the literature. For example, the tree representation of Nested Logit models is obviously a network, where each node has exactly one predecessor and the α parameters associated with the edges are all 1. Clearly, an arbitrary (finite) number of levels can be used. The Multinomial Logit Model being a special case of the Nested Logit, it also fits in this framework. The Cross-Nested model is a GEV model generated by

$$G(x_1, \ldots, x_J) = \sum_m \left(\sum_{j \in C} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}. \tag{38}$$

Eq. (38) has been proposed by Ben-Akiva and Bierlaire (1999). The derivation of the model and the elasticity computation have been described by Wen and Koppelman (2001), with a slightly different model formulation. They call it the *Generalised Nested Logit* model.

This model can also be represented in our framework. The network is composed of a root v_0 (associated with a parameter μ), a list of nodes $w_1, \ldots, w_M$ for the nests (with parameters μ_m such that $\mu \leq \mu_m$, $m = 1, \ldots, M$) and a list of nodes for the alternatives $v_1, \ldots, v_J$ (with parameters

μ_j, $j = 1, \ldots, J$). There is an edge between the root and each nest w_m, with a parameter $\alpha_{0m} = 1$, and an edge between each nest w_m and each alternative v_i, with a parameter α_{im}. We assume that $\mu_m \leq \mu_i$ if $\alpha_{im} \neq 0$. A simple example is illustrated in Figure 1. We derive the associated Network GEV model using the GEV-inheritance formulae.

From (21), the G function associated with each alternative is

$$G^i(x) = x_i^{\mu_i}, \tag{39}$$

the expected maximum utility is

$$\bar{U}_i = V_i + \frac{\gamma}{\mu_i}, \tag{40}$$

and the probability is

$$P_i(k) = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

The GEV-function G^m associated with each nest is given by

$$G^m(x) = \sum_{j \in \mathcal{C}} \alpha_{jm} (G^j(x))^{\frac{\mu_m}{\mu_j}} = \sum_{j \in \mathcal{C}} \alpha_{jm} x_j^{\mu_m}. \tag{42}$$

The expected maximum utility is

$$\bar{U}_m = \frac{\gamma}{\mu_m} + \frac{1}{\mu_m} \log \sum_{j \in \mathcal{C}} \alpha_{jm} e^{\mu_m V_j}. \tag{43}$$

The probability of choosing alternative i given by the model associated with node m is

$$P_m(i) = \frac{\alpha_{im} e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}} \alpha_{jm} e^{\mu_m V_j}}. \tag{44}$$

Finally, the GEV-function associated with the root is

$$G(x) = \sum_{m=1}^{M} (G^m(x))^{\frac{\mu}{\mu_m}} = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{45}$$

which is exactly (38). The expected maximum utility is

$$\bar{U} = \frac{\gamma}{\mu} + \frac{1}{\mu} \log \sum_{m=1}^{M} e^{\mu\left(\bar{U}_m - \frac{\gamma}{\mu_m}\right)}. \tag{46}$$

From (46), we have

$$\mu\left(\bar{U}_m - \frac{\gamma}{\mu_m}\right) = \log\left(\sum_{i\in\mathcal{C}} \alpha_{im} e^{\mu_m V_i}\right)^{\frac{\mu}{\mu_m}}, \tag{47}$$

and we obtain

$$\bar{U} = \frac{\gamma}{\mu} + \frac{1}{\mu} \sum_{m=1}^{M} \left(\sum_{i\in\mathcal{C}} \alpha_{im} e^{\mu_m V_i}\right)^{\frac{\mu}{\mu_m}}. \tag{48}$$

The probability of choosing a nest m is

$$P(m) = \frac{e^{\mu\left(\bar{U}_m - \frac{\gamma}{\mu_m}\right)}}{\sum_n e^{\mu\left(\bar{U}_n - \frac{\gamma}{\mu_n}\right)}} = \frac{\left(\sum_{i\in\mathcal{C}} \alpha_{im} e^{\mu_m V_i}\right)^{\frac{\mu}{\mu_m}}}{\sum_n \left(\sum_{i\in\mathcal{C}} \alpha_{in} e^{\mu_n V_i}\right)^{\frac{\mu}{\mu_n}}}, \tag{49}$$

where the second equality derives from (47).

Finally, the probability of choosing an alternative k is

$$P(k) = \sum_{m=1}^{M} P(m) P_m(k) = \frac{\left(\sum_{i\in\mathcal{C}} \alpha_{im} e^{\mu_m V_i}\right)^{\frac{\mu}{\mu_m}}}{\sum_n \left(\sum_{i\in\mathcal{C}} \alpha_{in} e^{\mu_n V_i}\right)^{\frac{\mu}{\mu_n}}} \frac{\alpha_{km} e^{\mu_m V_k}}{\sum_{j\in\mathcal{C}} \alpha_{jm} e^{\mu_m V_j}}. \tag{50}$$

The Nested-Logit formula can be derived from the above equations by setting $\alpha_{im} = 1$ if alternative i belongs to nest m, and 0 otherwise.

Of course, this applies automatically to all special instances of the CNL, such as the Paired Combinatorial Logit model (Koppelman and Wen, 2000), the Generalised Nested Logit model (Wen and Koppelman, 2001), the Ordered GEV model (Small, 1987), the Link-Nested Logit model (Vovsha and Bekhor, 1998), the GenL model (Swait, 2001) or specialised compound GEV models (Bhat, 1998, Whelan et al., 2002).

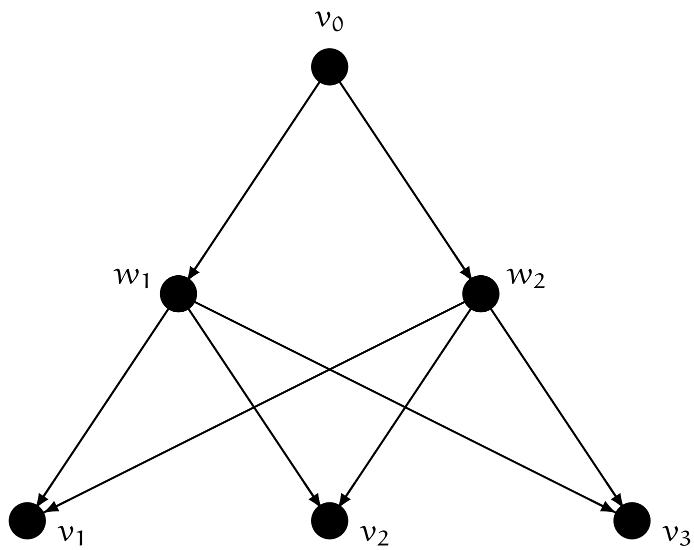

Figure 1: A simple network

7 Estimation

The model inherits the closed form advantage of the GEV family. Efficient nonlinear programming techniques can therefore be used to identify the maximum likelihood estimator of the parameters. To achieve this goal, the analytical derivatives of the log-likelihood function with respect to each unknown parameter are required. They are provided in Appendix B.

The freeware package BIOGEME (Bierlaire, 2003b, Bierlaire, 2003a), available at

`http://roso.epfl.ch/biogeme`

allows the estimation of Network GEV models. As input, the user must define the vertices and edges of the network, and choose which parameters are fixed and which must be estimated.

An alternative approach to estimating the model is to note that it can be re-formulated as a tree logit model with each path from root to alternative being represented as an alternative in the tree logit model (Daly, 1987, Daly, 2001a). Then the choice probability given by the network model is identical to the probability of choice of any of a set of alternatives in a tree (nested) logit model. This transformation means that the Network GEV model can also be estimated by a suitable adaptation of software designed for the estimation of tree logit models.

It is important to note that the estimation process may happen to force some parameters α_{ij} to 0 and, consequently, to produce a model that violates the assumptions of the GEV Network (see Section 4). In this case, all arcs (i, j) such that $\alpha_{ij} = 0$ must be removed from the model, in order to obtain a valid Network GEV model. It is then the analyst's responsibility to decide if the resulting model still captures the choice situation under analysis.

The Network representation generates models with many parameters. Indeed, there is one parameter per node in the network and one parameter per arc. Not all of them are identifiable from data. A thorough analysis of the identifiability issues has not yet been performed. Some preliminary comments can be made though. Firstly, the homogeneity parameters are relevant only in terms of their ratio, exactly like for Nested Logit models. Therefore, a similar normalisation (from the top or the bottom) is required. Also, not all parameters α associated with the edges of the network can be identified. Indeed, multiplying all parameters α in (23) by a constant is equivalent to multiply the G^i function by the constant, resulting in a modification of the scale of the utilities (due to the homogeneity of G^i) which does not affect the probability. A normalisation is also required there.

8 Conclusions and future research

The two main contributions of this paper are the GEV inheritance theorems and the GEV Network representation. They allow the design of GEV models based on intuitive interpretation of the application, similar to the development of trees for Nested Logit models, with no additional theoretical analysis required from the analyst. The Multinomial logit, the Nested logit and the Cross Nested logit models can all be represented by a GEV Network. The probability, expected maximum utility and elasticity formula have been derived. In the special case of the Cross-Nested Logit model, our results match those presented by Wen and Koppelman (2001).

The framework we propose is operational in two ways. Firstly, the model specification is simplified thanks to the network structure. Secondly, the estimation of the models is possible with BIOGEME and may become possible with other packages. We believe that the results proposed in this paper will provide more research opportunities, as the investigation of new GEV models will not be subject any more to heavy mathematical proofs.

The RNEV or Network GEV model is more general than any GEV model presented in previous literature known to the authors. It allows the analyst the freedom to construct networks consistent with an intuitive view of the structure of the choice situation being represented and should form an attractive basis for applications and the development of further theoretical understanding of GEV framework.

A Detailed proof for Theorem 4

In this Appendix, we extend the proof of GEV-differentiability in Theorem 4 by explicitly using the exact derivatives of the function. We start with two technical lemmas[1]. The first lemma provides a formula for the derivative of the function G^β.

Lemma 14 *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*

$$S^p(x) = G(x)^{\beta-p} \prod_{i=0}^{p-1} (\beta - i). \tag{51}$$

Given a set R containing r indices, we define

$$D_R = \frac{\partial^r G(x)}{(\partial x_i)_{i \in R}}, \tag{52}$$

similarly to (3). Then, we have

$$\frac{\partial^k}{\partial x_1 \dots \partial x_k} G(x)^\beta = \sum_{P \in \mathcal{P}_k} S^p \prod_{R \in P} D_R. \tag{53}$$

Proof. The proof is by induction. The cases $k = 1, 2, 3$ are obtained by simple calculus.

The case $k = 1$ is derived in (12). For the second derivative, we have

$$\begin{aligned} D^*_{\{1,2\}}(x) &= \tfrac{\partial^2}{\partial x_1 \partial x_2} G(x)^\beta \\ &= \beta(\beta-1) G(x)^{\beta-2} \tfrac{\partial G}{\partial x_1}(x) \tfrac{\partial G}{\partial x_2}(x) + \beta G(x)^{\beta-1} \tfrac{\partial^2 G}{\partial x_1 \partial x_2}(x) \\ &= \beta(\beta-1) G(x)^{\beta-2} G_{\{1\}}(x) G_{\{2\}}(x) + \beta G(x)^{\beta-1} G_{\{1,2\}}(x). \end{aligned} \tag{54}$$

For $k = 3$, we have

$$\begin{aligned} \tfrac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} &= \beta(\beta-1)(\beta-2) G(x)^{\beta-3} \tfrac{\partial G}{\partial x_1} \tfrac{\partial G}{\partial x_2} \tfrac{\partial G}{\partial x_3} \\ &+ \beta(\beta-1) G(x)^{\beta-2} \left(\tfrac{\partial G}{\partial x_1} \tfrac{\partial^2 G}{\partial x_2 \partial x_3} + \tfrac{\partial G}{\partial x_2} \tfrac{\partial^2 G}{\partial x_1 \partial x_3} + \tfrac{\partial G}{\partial x_3} \tfrac{\partial^2 G}{\partial x_1 \partial x_2} \right) \\ &+ \beta G(x)^{\beta-1} \tfrac{\partial^3 G}{\partial x_1 \partial x_2 \partial x_3} \end{aligned} \tag{55}$$

[1]The authors would like to thank Jean-Albert Ferrez and Alain Prodon for their help with set partitioning.

We assume now that the result is true for k, and we prove it is true also for $k+1$. We first compute

$$\frac{\partial^{k+1}}{\partial x_1 \dots \partial x_{k+1}} G(x)^\beta = \frac{\partial}{\partial x_{k+1}} \left(\frac{\partial^k}{\partial x_1 \dots \partial x_k} G(x)^\beta \right) \tag{56}$$

that is

$$\begin{aligned} & \frac{\partial^{k+1}}{\partial x_1 \dots \partial x_{k+1}} G(x)^\beta \\ = & \frac{\partial}{\partial x_{k+1}} \left(\sum_{P \in \mathcal{P}_k} S^p \prod_{R \in P} D_R \right) \\ = & \sum_{P \in \mathcal{P}_k} \left[\left(\frac{\partial}{\partial x_{k+1}} S^p \right) \prod_{R \in P} D_R + S^p \frac{\partial}{\partial x_{k+1}} \prod_{R \in P} D_R \right], \end{aligned} \tag{57}$$

where

$$\begin{aligned} \frac{\partial}{\partial x_{k+1}} S^p &= \prod_{i=0}^{p-1} (\beta - i)(\beta - p) G(x)^{\beta-(p+1)} \frac{\partial G}{\partial x_{k+1}} \\ &= S^{p+1} \frac{\partial G}{\partial x_{k+1}} \end{aligned} \tag{58}$$

and

$$\frac{\partial}{\partial x_{k+1}} \prod_{R \in P} D_R = \sum_{R \in P} \frac{\partial}{\partial x_{k+1}} D_R \prod_{T \in P, T \neq R} D_T. \tag{59}$$

Consequently,

$$\begin{gathered} \frac{\partial^{k+1}}{\partial x_1 \dots \partial x_{k+1}} G(x)^\beta = \\ \sum_{P \in \mathcal{P}_k} \left[\left(S^{p+1} \frac{\partial G}{\partial x_{k+1}} \right) \prod_{R \in P} D_R + S^p \sum_{R \in P} D_{R \cup \{k+1\}} \prod_{T \in P, T \neq R} D_T \right]. \end{gathered} \tag{60}$$

Then, we prove the result by obtaining (60) directly from (53). Using

(69) in (53), we have

$$\frac{\partial^{k+1}}{\partial x_1 \ldots \partial x_{k+1}} G(x)^\beta =$$

$$\sum_{i=1}^{p} \left[S^{p+1} \prod_{R \in P \cup \{k+1\}} D_R + \sum_{\ell=1}^{n_i} S^p \prod_{R \in P_{i,\ell}^{k+1}} D_R \right] \tag{61}$$

$$\sum_{i=1}^{p} \left[S^{p+1} \frac{\partial G}{\partial x_{k+1}} \prod_{R \in P} D_R + S^p \sum_{\ell=1}^{n_i} \prod_{R \in P_{i,\ell}^{k+1}} D_R \right].$$

We finally use the definition (70) of $P_{i,\ell}^{k+1}$ to obtain (60) and prove the result. □

The second lemma analyses the sign of each term of (53).

Lemma 15 *For an arbitrary* k*, we consider one term of (53) with* $0 < \beta \leq 1$*. Let* P *be a partition belonging to* $\mathcal{P}_k$*, composed of* p *sets of indices,* $p \geq 1$*. Then, the sign of* $S^p(x) \prod_{R \in P} D_R$ *and the sign of its derivative*

$$\frac{\partial}{\partial x_{k+1}} S^p(x) \prod_{R \in P} D_R \tag{62}$$

are opposite.

Proof. We have

$$\frac{\partial}{\partial x_{k+1}} S^p(x) \prod_{R \in P} D_R \tag{63}$$

$$= \left(\frac{\partial}{\partial x_{k+1}} S^p(x) \right) \prod_{R \in P} D_R \tag{64}$$

$$+ \quad S^p(x) \frac{\partial}{\partial x_{k+1}} \prod_{R \in P} D_R. \tag{65}$$

The sign of (58) is entirely determined by $\prod_{i=0}^{p-1} (\beta - i)(\beta - p)$, as both $G(x)$ and $\partial G / \partial x_{k+1}$ are non-negative. From (51), the sign of S^p is entirely determined by $\prod_{i=0}^{p-1}(\beta-i)$. As $\beta-p \leq 0$, the sign of (64) is consequently opposite to the sign of $S^p \prod_{R\in P} D_R$.

The modification of the sign of (65) is a direct consequence of the definition (52) of D_R and property 4 of GEV models. $\square$

We are now able to provide a formal proof for item 4 of Theorem 4. Indeed, from Lemma 14, we have

$$\frac{\partial^k}{\partial x_1 \dots \partial x_k} G(x)^\beta = \sum_{P\in\mathcal{P}_k} S^p \prod_{R\in P} D_R, \tag{66}$$

where $S^p(x)$ is defined by (51) and D_R is defined by (52). The sign alternance is proved recursively. For $k = 1$, the derivative (12) is trivially non-negative. For $k = 2$, the derivative (54) is non-positive, as $\beta \leq 1$, and $\partial^2 G/\partial x_1 \partial x_2$ is non-positive from property 4 of GEV models. Then, Lemma 15 is applied for further iterates of the recursion, showing that G^β is GEV-differentiable.

B Derivatives of the model

In this section, we provide the derivatives of the model (16) with regard to specific parameters, which must be provided to optimisation packages. The derivative of G with regard to a variable x_k is

$$\frac{\partial G^i(x)}{\partial x_k} = \sum_j \alpha_{ij} \frac{\mu_i}{\mu_j} G^j(x)^{\frac{\mu_i}{\mu_j}-1} \frac{\partial G^j(x)}{\partial x_k}.$$

The derivative with regard to α_{ij} is simply

$$\frac{\partial G^i(x)}{\alpha_{ij}} = G^j(x)^{\frac{\mu_i}{\mu_j}},$$

and the derivative with regard to μ_i

$$\frac{\partial G^i(x)}{\partial \mu_i} = \sum_j \frac{\alpha_{ij}}{\mu_j} G^j(x)^{\frac{\mu_i}{\mu_j}} \log G^j(x).$$

If node k is a successor of node i, then

$$\frac{\partial G^i(x)}{\mu_k} = \alpha_{ik} G^k(x)^{\frac{\mu_i}{\mu_k}-1} \frac{\mu_i}{\mu_k} \left(\frac{\partial G^k(x)}{\partial \mu_k} - \frac{1}{\mu_k} G^k(x) \log G^k(x) \right)$$

If not,

$$\frac{\partial G^i(x)}{\mu_k} = \sum_j \alpha_{ij} \frac{\mu_i}{\mu_j} G^j(x)^{\frac{\mu_i}{\mu_j}-1} \frac{\partial G^j(x)}{\partial \mu_k}.$$

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

$$\mathcal{P}_k = \bigcup_{i=1}^{p} P_i^k \tag{67}$$

where

$$P_i^k = \bigcup_{j=1}^{n_i} R_j^i \tag{68}$$

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

$$\mathcal{P}_{k+1} = \bigcup_{i=1}^{p} \left[\left(P_i^k \cup \{k+1\} \right) \bigcup_{\ell=1}^{n_i} P_{i,\ell}^{k+1} \right], \tag{69}$$

where

$$P_{i,\ell}^{k+1} = \{R_\ell^i \cup (k+1)\} \cup \bigcup_{\substack{j=1 \\ j \neq \ell}}^{n_i} R_j^i. \tag{70}$$

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

$$\begin{aligned} \mathcal{P}_2 &= \textstyle\bigcup_{i=1}^1 \left[\left(P_i^1 \cup \{2\} \right) \bigcup_{\ell=1}^1 P_{i,\ell}^2 \right], \\ &= P_1^1 \cup \{2\} \cup P_{1,1}^2 \\ &= \{\{1\}\{2\}\} \cup P_{1,1}^2 \end{aligned} \tag{71}$$

where

$$P^2_{1,1} = \{R^1_1 \cup 2\} \cup \bigcup_{\substack{j=1\\j\neq 1}}^{1} R^1_j = \{\{1,2\}\}. \tag{72}$$

Therefore, $\mathcal{P}_2 = \{\{\{1\}\{2\}\}, \{\{1,2\}\}\}$. Denoting $P^2_1 = \{R^1_1, R^1_2\}$, with $R^1_1 = \{1\}$ and $R^1_2 = \{2\}$ and $P^2_2 = \{R^2_1\}$,where $R^2_1 = \{1,2\}$, we can compute $\mathcal{P}_3$.

$$\begin{aligned} \mathcal{P}_3 &= \textstyle\bigcup_{i=1}^{2} \left[\left(P^2_i \cup \{3\}\right) \bigcup_{\ell=1}^{n_i} P^3_{i,\ell} \right], \\ &= \left[\left(P^2_1 \cup \{3\}\right) \textstyle\bigcup_{\ell=1}^{2} P^3_{1,\ell} \right] \bigcup \left[\left(P^2_2 \cup \{3\}\right) \textstyle\bigcup_{\ell=1}^{1} P^3_{2,\ell} \right] \\ &= \left[\left(P^2_1 \cup \{3\}\right) \cup P^3_{1,1} \cup P^3_{1,2} \right] \bigcup \left[\left(P^2_2 \cup \{3\}\right) \cup P^3_{2,1} \right] \end{aligned} \tag{73}$$

where

$$\begin{aligned} P^3_{1,1} &= \{R^1_1 \cup 3\} \cup \textstyle\bigcup_{\substack{j=1\\j\neq 1}}^{2} R^1_j \\ &= \{R^1_1 \cup 3\} \cup R^1_2 \\ &= \{\{1\} \cup 3\} \cup \{2\}, \\ &= \{\{1,3\}\{2\}\} \\ P^3_{1,2} &= \{\{2,3\}\{1\}\} \\ P^3_{2,1} &= \{\{1,2,3\}\} \end{aligned} \tag{74}$$

Consequently,

$$\begin{gathered} \mathcal{P}_3 = \{ \\ \{\{1\}\{2\}\{3\}\}, \\ \{\{1,3\}\{2\}\}, \\ \{\{2,3\}\{1\}\}, \\ \{\{1,2\}\{3\}\}, \\ \{\{1,2,3\}\} \\ \} \end{gathered} \tag{75}$$

D Some results

D.1 Derivation of (4)

Euler's formula states that

$$\mu G = \sum_j y_j G_j.$$

Therefore, (1) reads

$$\frac{y_i G_i}{\sum_j y_j G_j}. \tag{76}$$

To obtain (4), we note that

$$y_i G_i = e^{V_i} G_i = e^{\log(e^{V_i} G_i)} = e^{V_i + \log G_i}.$$

D.2 Derivation of (5) and (6)

We rewrite here McFadden's proof (p. 82). Using the μ-homogeneity of G, we have

$$\begin{aligned} F_i(\langle V_i + \varepsilon_i - V_j \rangle) &= e^{-a e^{-\mu(V_i+\varepsilon_i)}} e^{(-V_i-\varepsilon_i)(\mu-1)} G_i(\langle e^{V_j} \rangle) e^{-\varepsilon_i} \\ &= e^{-a e^{-\mu w}} e^{-w(\mu-1)} G_i(\langle e^{V_j} \rangle) e^{V_i} e^{-w} \end{aligned} \tag{77}$$

and Eq. (16) becomes

$$\begin{aligned} \bar{U} &= \sum_i \int_w w e^{-a e^{-\mu w}} e^{-w(\mu-1)} G_i(\langle e^{V_j} \rangle) e^{V_i} e^{-w} dw \\ &= \int_w w e^{-a e^{-\mu w}} e^{-w(\mu-1)} \mu a e^{-w} dw \\ &= \int_w w e^{-a e^{-\mu w}} \mu a e^{-\mu w} dw \end{aligned} \tag{78}$$

Let $t = \mu w$ and $dt = \mu dw$ to obtain

$$\bar{U} = \frac{1}{\mu} \int_t t e^{-a e^{-t}} a e^{-t} dt = \frac{\log G(x) + \gamma}{\mu}. \tag{79}$$

Eq. (6) results directly from (1).

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.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D., Daly, A., de Palma, A., Gopinath, D., Karlstrom, A. and Munizaga, M. A. (2002). Hybrid choice models: Progress and challenges, *Marketing Letters* . accepted for publication.

Bhat, C. (1998). Analysis of travel mode and departure time choice for urban shopping trips, *Transportation Research B* **32**(6): 361–371.

Bierlaire, M. (2001). A general formulation of the cross-nested logit model, *Proceedings of the 1st Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. (2002). The network GEV model, *Proceedings of the 2nd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. (2003a). BIOGEME: a free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. (2003b). An introduction to BIOGEME version 0.6. roso.epfl.ch/biogeme.

Dagsvik, J. (1994). Discrete and continuous choice, max-stable process, and independence from irrelevant alternatives, *Econometrica* **62**: 1179–1205.

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

Daly, A. (2001a). Alternative tree logit models: comments on a paper of Koppelman and Wen, *Transportation Research B* **35**(8): 717–724.

Daly, A. (2001b). Recursive nested EV model, *ITS Working Paper 559*, Institute for Transport Studies, University of Leeds.

Daly, A. and Bierlaire, M. (2003). A general and operational representation of GEV models, *Technical Report RO-030502*, Institute of

MAthematics, Operations Research Group ROSO, EPFL, Lausanne, Switzerland.

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

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.

McFadden, D. (1981). Econometric models of probabilistic choice, *in* C. Manski and D. McFadden (eds), *Structural analysis of discrete data with econometric application*, MIT Press, Cambridge, Mass.

McFadden, D. L. (2000). Disaggregate behavioral travel demand's RUM side: a 30-year retrospective, *International Association for Travel Behavior Confenrece*, Gold Coast, Quennsland.

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

Prashker, J. N. and Bekhor, S. (2000). Stochastic user-equilibrium formulations for extended-logit assignment models, *Transportation Research Record* **1676**. paper 99-0645.

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, *Transportation Research B* **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.

Whelan, G., Batley, R., Fowkes, T. and Daly, A. (2002). Flexible models for analysing route and departure time choice, Presented at European Transport Conference, Cambridge.