Ann Oper Res (2006) 144:287–300
DOI 10.1007/s10479-006-0015-x

A theoretical analysis of the cross-nested logit model

Michel Bierlaire

Published online: 25 May 2006

Abstract The emergence of Intelligent Transportation Systems and the associated technologies has increased the need for complex models and algorithms. Namely, real-time information systems, directly influencing transportation demand, must be supported by detailed behavioral models capturing travel and driving decisions. Discrete choice models methodology provide an appropriate framework to capture such behavior. Recently, the Cross-Nested Logit (CNL) model has received quite a bit of attention in the literature to capture decisions such as mode choice, departure time choice and route choice.

In this paper, we develop on the general formulation of the Cross Nested Logit model proposed by Ben-Akiva and Bierlaire (1999) and based on the Generalized Extreme Value (GEV) model. We show that it is equivalent to the formulations by Papola (2004) and Wen and Koppelman (2001). We also show that the formulations by Small (1987) and Vovsha (1997) are special cases of this formulation. We formally prove that the Cross-Nested Logit model is indeed a member of the GEV models family. In doing so, we clearly distinguish between conditions that are necessary to prove consistency with the GEV theory, from normalization conditions. Finally, we propose to estimate the model with non-linear programming algorithms, instead of heuristics proposed in the literature. In order to make it operational, we provide the first derivatives of the log-likelihood function, which are necessary to such optimization procedures.

Keywords Transportation demand · Behavior model · Logit · GEV · Random utility

1. Introduction

The emergence of Intelligent Transportation Systems and the associated technologies has increased the need for complex models and algorithms. Namely, real-time information systems, directly influencing transportation demand, must be supported by detailed behavioral

M. Bierlaire (✉)
Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne,
Station 8 CH-1015 Lausanne, Switzerland
e-mail: michel.bierlaire@epfl.ch

models capturing travel and driving decisions (Bierlaire, Mishalani and Ben-Akiva, 2000; Ben-Akiva et al., 2001; Chatterjee et al., 2002).

Discrete choice models methodology provide an appropriate framework to capture such behavior. 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 particular structure of transportation related choice situations is not always fully consistent with the underlying modeling theory (Ben-Akiva and Bierlaire, 2003), requiring to enhance and adapt existing models. The GEV models family has been proved to be consistent with random utility theory by McFadden (1978). It appears that only a few members of this family have been exploited so far. In addition to the well known multinomial logit and nested logit models, the *cross-nested logit model* has recently received some attention in the literature, although it has already been mentioned by McFadden (1978). It is a natural extension of the nested logit model, where each alternative can potentially belong to more than one nest, allowing for a more complex correlation structure.

The name *cross-nested* seems to be due to Vovsha (1997), who uses this model for a mode choice survey in Israel. Vovsha's model is similar to the Ordered GEV model proposed by Small (1987). This model is appealing for its ability to capture a wide variety of correlation structures. Papola (2004) has conjectured that a specific CNL model can be obtained for any given homoscedastic variance-covariance matrix, but Abbé, Bierlaire and Toledo (2005) have shown this results not to hold in general. The CNL model has a closed form formulation derived from the GEV model. Therefore, it is appropriate for a wide range of applications. Vovsha (1997) and Bierlaire, Axhausen and Abbay (2001) use a CNL model for mode choice. It has also been shown to be appropriate for route choice applications (Vovsha and Bekhor, 1998), where topological correlations cannot be captured correctly by the multinomial and the nested logit models. Namely, Prashker and Bekhor (1999) discuss the use of route choice models based on a simplified CNL model within the stochastic user equilibrium context. Swait (2001) suggests an original CNL structure to model the choice set generation process.

As part of the GEV model family, the Cross-Nested Logit model inherits the homoscedastic property. However, heteroscedastic versions of the model can easily be derived (see, for instance, Bhat, 1995; Zeng, 2000).

The Cross-Nested model is appealing for modeling complex choice situations because

- it inherits from the GEV family the theoretical foundations of random utility theory,
- it inherits from the GEV family the closed form of the probability model,
- it allows to capture a wide range of correlation structures,
- the Multinomial logit and the Nested logit models are special cases of the Cross-Nested logit model.

A detailed theoretical analysis of the model is therefore necessary. The most thorough analysis of the CNL model is probably due to Wen and Koppelman (2001), who present it as the *Generalized Nested Logit Model*. They show how other models are specific cases of that model, and provide direct and cross elasticities of probabilities with respect to changes in attributes. Wen and Koppelman (2001) state that the model is indeed a GEV model, without actually proving it. In this paper, we provide a formal proof, and clearly identify the conditions associated with the model validity. The other objective of the paper is to suggest an estimation procedure, based on classical non-linear programming techniques applied to maximum likelihood estimation, instead of heuristics presented in the literature.

Such techniques require derivatives of the log-likelihood function, which are provided in the appendix.

Section 2 introduces the GEV model and presents various formulations of the Cross-Nested Logit model from the literature. In Section 3, we analyze the most general formulation. We prove that it is consistent with the GEV model family. The estimation procedure is discussed in Section 4. In the appendix, we provide the derivatives of the model with regard to parameters to be estimated. Those are required for most efficient optimization algorithms.

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 and the Nested Logit models. The probability of choosing alternative i within the choice set C of a given choice maker is

$$P(i \mid C) = \frac{x_i \frac{\partial G}{\partial x_i}(x_1, \ldots, x_J)}{\mu G(x_1, \ldots, x_J)} \tag{1}$$

where J is the number of available alternatives, $x_i = e^{V_i}$, V_i is the deterministic part of the utility function associated with 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 $G(\alpha x) = \alpha^\mu G(x)$,
2. $\lim_{x_i \to +\infty} G(x_1, \ldots, x_i, \ldots, x_J) = +\infty$, for each $i = 1, \ldots, J$,
3. the kth partial derivative with respect to k distinct x_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

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

Because G is homogeneous, Euler's formula implies that

$$\mu G = \sum_j x_j G_j, \tag{3}$$

where $G_i = \frac{\partial G}{\partial x_i}$. Therefore, (1) can be written

$$\frac{x_i G_i}{\sum_j x_j G_j}. \tag{4}$$

Given that

$$x_i G_i = e^{V_i} G_i = e^{\ln(e^{V_i} G_i)} = e^{V_i + \ln G_i}, \tag{5}$$

we obtain a nice form for the probability model:

$$P(i \mid C) = \frac{e^{V_i + \ln G_i(\ldots)}}{\sum_{j=1}^{J} e^{V_j + \ln G_j(\ldots)}}. \tag{6}$$

It is well known that the Multinomial Logit and the Nested Logit models are instances of this model family, with

$$G(x) = \sum_{j \in C} x_j^{\mu} \tag{7}$$

for the Multinomial Logit and

$$G(x) = \sum_{m=1}^{M} \left(\sum_{i=1}^{J_m} x_i^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \tag{8}$$

for the Nested Logit model with M nests containing J_m alternatives each. We present now several formulations of the Cross-Nested Logit model proposed in the literature.

2.1. Formulations of the cross-nested logit model

The limitations of the Nested Logit model have been observed by several authors (Williams, 1977; Forinash and Koppelman, 1993). The requirement of unambiguous assignment of alternatives to nests does not allow to capture mixed interactions across alternatives. We present here some formulations proposed in the literature, adopting the notations of the respective authors.

After McFadden (1978) seminal paper, it seems that the first Cross-Nested Logit model has been proposed by Small (1987) in the context of departure time choice. Small's model, called the *Ordered GEV model*, is based on the following function:

$$G(x_1, \ldots, x_J) = \sum_{r=1}^{J+M} \left(\sum_{j \in B_r} w_{r-j} x_j^{1/\rho_r} \right)^{\rho_r}, \tag{9}$$

where M is a positive integer, ρ_r and w_m are constants satisfying $0 < \rho_r \leq 1$, $w_m \geq 0$ and

$$\sum_{m=0}^{M} w_m = 1. \tag{10}$$

The B_r are overlapping subsets of alternatives:

$$B_r = \{j \in \{1, \ldots, J\} | r - M \leq j \leq r\}. \tag{11}$$

Vovsha (1997) introduces the name "Cross-Nested Logit", and applies the model to a mode choice application, where the "park & ride" alternative is allowed to belong to the "composite auto" and the "composite transit" nests. Vovsha derives the Cross-Nested Logit model from the GEV model with the generating function:

$$G(x_1, \ldots, x_J) = \sum_m \left(\sum_{j \in C} \alpha_{jm} x_j \right)^{\mu} \tag{12}$$

where m is the nest index, and α_{jm} are model parameters such that

$$0 \leq \alpha_{jm} \leq 1 \quad \forall j, m, \tag{13}$$

and

$$\sum_m \alpha_{jm} > 0 \quad \forall j. \tag{14}$$

Vovsha (1997) imposes also that

$$\sum_m \alpha_{jm}^{\mu} = 1 \quad \forall j. \tag{15}$$

Ben-Akiva and Bierlaire (1999) mention the CNL as an example of an instance of a GEV model based on the following generating function:

$$G(x_1, \ldots, x_J) = \sum_m \left(\sum_{j \in C} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{16}$$

where m is the nest index, and μ_m is a parameter associated with nest m.

A similar formulation is used by Papola (2004), based on the following generating function:

$$G(x_1, \ldots, x_J) = \sum_k \left(\sum_{j \in C_k} \alpha_{jk}^{\theta_0/\theta_k} e^{V_j/\theta_k} \right)^{\frac{\theta_k}{\theta_0}}, \tag{17}$$

where $C_k \subseteq C$ is the set of alternatives in nest k, and $0 \leq \theta_k \leq \theta_0$. Papola imposes also that

$$\sum_k \alpha_{jk} = 1 \quad \forall j. \tag{18}$$

Finally, Wen and Koppelman (2001) also provide an analysis of the CNL model, naming it the *Generalized Nested Logit Model* based on the following generating function:

$$G(x_1, \ldots, x_J) = \sum_m \left(\sum_{n' \in N_m} (\alpha_{n'm} x_{n'})^{\frac{1}{\mu_m}} \right)^{\mu_m}, \tag{19}$$

where $\alpha_{n'm} \geq 0$ and $0 < \mu_m \leq 1$, and N_m is the set of alternatives in nest m. The condition

$$\sum_m \alpha_{n'm} = 1 \quad \forall n' \tag{20}$$

is mentioned to provide a useful interpretation of the nest allocation. Also, Wen and Koppelman (2001) provide direct- and cross-elasticities formulae for the model.

Note that (12) and (16) allow all alternatives to belong to all nests, whereas (17) and (19) explicitly define the set of alternatives within each nest ($\mathcal{C}_k$ and N_m, resp.). This makes no difference if we define $\alpha_{im} = 0$ if and only if alternative i does not belong to nest m.

There is a trend in the discrete choice community to use the name *cross-nested* only when the parameters capturing the level of membership to nests (usually denoted by α) are not estimated but imposed a priori. We prefer to use it in the general case.

3. Theoretical analysis

Among these formulations, (16) is the most general. Indeed, Vovsha's and Small's formulations are specific cases of (16). We obtain Small's formulation (9) with $\mu = 1$ and $\mu_m = 1/\rho_m$. Vovsha's formulation (12) is obtained from (16) with $\mu_m = 1$ for all m.

Papola's model (17) is equivalent to (16), with $\mu = 1/\theta_0$, $\mu_m = 1/\theta_m$ and $\alpha_{jm} = \alpha_{jm}^{\theta_0/\theta_m}$.

Wen and Koppelman's model (19) is equivalent to (16) with $\mu = 1$, which is a common normalization for GEV models.

No formal proof is given in the literature that the CNL model is indeed a GEV model. In most papers, a proof is sketched, but condition 3 is never derived completely. Theorem 2 shows that (16) is indeed a GEV generating function, and identifies the sufficient conditions on the parameters. Its proof is based on the following lemma.

Lemma 1. *Let $i_1, \ldots, i_k$ be k different indices ($k > 0$) chosen within $\{1, \ldots, J\}$. If G is defined by (16), then*

$$\partial^k G(x)/\partial x_{i_1} \ldots \partial x_{i_k} = \sum_m \left(\mu_m^k \prod_{n \in \{i_1, \ldots, i_k\}} \left(\alpha_{nm} x_n^{\mu_m - 1}\right) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) y_m^{\frac{\mu - k\mu_m}{\mu_m}} \right) \tag{21}$$

where

$$y_m = \sum_{j \in C} \alpha_{jm} x_j^{\mu_m}. \tag{22}$$

Proof: The proof is by induction. We have

$$\begin{aligned} \frac{\partial^k G(x)}{\partial x_{i_1}} &= \sum_m \left(\frac{\mu}{\mu_m} y_m^{\frac{\mu - \mu_m}{\mu_m}} \mu_m \alpha_{i_1 m} x_{i_1}^{\mu_m - 1} \right) \\ &= \sum_m \left(\mu_m \alpha_{i_1 m} x_{i_1}^{\mu_m - 1} \frac{\mu}{\mu_m} y_m^{\frac{\mu - \mu_m}{\mu_m}} \right) \end{aligned}$$

proving the result for $k = 1$.

Assuming now that the result is verified for k, we have

$$
\begin{aligned}
&\partial^{k+1} G(x)/\partial x_{i_1} \ldots \partial x_{i_{k+1}} \\
&= \frac{\partial}{\partial x_{i_{k+1}}} \frac{\partial^k G(x)}{\partial x_{i_1} \ldots \partial x_{i_k}} \\
&= \frac{\partial}{\partial x_{i_{k+1}}} \sum_m \left(\mu_m^k \prod_{n=i_1}^{i_k} \left(\alpha_{nm} x_n^{\mu_m - 1}\right) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) y_m^{\frac{\mu - k\mu_m}{\mu_m}} \right) \\
&= \sum_m \mu_m^k \prod_{n=i_1}^{i_k} \left(\alpha_{nm} x_n^{\mu_m - 1}\right) \prod_{n=0}^{k-1} \left(\frac{\mu}{\mu_m} - n \right) \left(\frac{\mu}{\mu_m} - k \right) y_m^{\frac{\mu - k\mu_m - \mu_m}{\mu_m}} \alpha_{i_{k+1}} \mu_m x_{i_{k+1}}^{\mu_m - 1} \\
&= \sum_m \left(\mu_m^{k+1} \prod_{n=i_1}^{i_{k+1}} \left(\alpha_{nm} x_n^{\mu_m - 1}\right) \prod_{n=0}^{k} \left(\frac{\mu}{\mu_m} - n \right) y_m^{\frac{\mu - (k+1)\mu_m}{\mu_m}} \right).
\end{aligned}
$$

That concludes the proof. □

Theorem 2. *The following conditions are sufficient for (16) to define a GEV generating function:*

1. $\alpha_{jm} \geq 0, \forall j, m,$
2. $\sum_m \alpha_{jm} > 0, \forall j,$
3. $\mu > 0,$
4. $\mu_m > 0, \forall m,$
5. $\mu \leq \mu_m, \forall m.$

Proof: We show that, under these assumptions, (16) verifies the four properties of GEV generating functions.

1. G is obviously non-negative, if $x \in \mathbb{R}^n_+$.
2. G is homogeneous of degree μ. Indeed,

$$
\begin{aligned}
G(\beta x) &= \sum_m \left(\sum_{j \in C} \alpha_{jm} \beta^{\mu_m} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \sum_m \left(\beta^{\mu_m} \sum_{j \in C} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \sum_m \beta^{\mu} \left(\sum_{j \in C} \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \\
&= \beta^{\mu} G(x).
\end{aligned}
$$

3. The limit properties hold from assumption 2, that guarantees that there is at least one non-zero coefficient α_{jm} for each alternative j.

$$\begin{aligned}\lim_{x_i\to\infty} G(x_1,\ldots,x_J) &= \lim_{x_i\to\infty}\sum_m\left(\sum_{j\in C}\alpha_{jm}x_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}}\\ &= \sum_m\left(\lim_{x_i\to\infty}\left(\sum_{j\in C}\alpha_{jm}x_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}}\right)\\ &= \infty\end{aligned}$$

4. The condition for the sign of the derivatives is obtained from Lemma 1.
Considering (21), we distinguish three cases, considering only $x \geq 0$.
   (a) If $k = 1$, we have

$$\partial G(x)/\partial x_j = \mu\sum_m\left(\alpha_{jm}x_j^{\mu_m-1}y_m^{\frac{\mu}{\mu_m}-1}\right) \geq 0. \tag{23}$$

   (b) If $k > 1$ and $\mu = \mu_m$, we have

$$\partial^k G(x)/\partial x_{i_1}\ldots\partial x_{i_k} = 0. \tag{24}$$

   Indeed,

$$\prod_{n=0}^{k-1}\left(\frac{\mu}{\mu_m} - n\right) \tag{25}$$

   contains a zero factor when $n = 1$.
   (c) If $k > 1$ and $\mu < \mu_m$, the sign of (21) is entirely determined by the sign of (25). For $n > 0$, we have $\frac{\mu}{\mu_m} - n < 0$ (assumption 5). Therefore, there are $k - 1$ negative and one positive factors in the product. We obtain that

$$\prod_{n=0}^{k-1}\left(\frac{\mu}{\mu_m} - n\right)\begin{cases}\geq 0 & \text{if } k \text{ is odd}\\ \leq 0 & \text{if } k \text{ is even}\end{cases} \tag{26}$$

   Therefore, in any case, we have

$$\partial^k G(x)/\partial x_{i_1}\ldots\partial x_{i_k}\begin{cases}\geq 0 & \text{if } k \text{ is odd}\\ \leq 0 & \text{if } k \text{ is even}\end{cases} \tag{27}$$

□

The probability formula can be directly derived from (1) and (16). From (23), we have

$$G_i = \mu\sum_m\alpha_{im}x_i^{\mu_m-1}\left(\sum_j\alpha_{jm}x_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}-1}.$$

Using (1), we obtain

$$P(i \mid \mathcal{C}) = \frac{\sum_m \alpha_{im} x_i^{\mu_m} \left(\sum_j \alpha_{jm} x_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}-1}}{\sum_n \left(\sum_j \alpha_{jn} x_j^{\mu_n}\right)^{\frac{\mu}{\mu_n}}}.$$

Re-arranging the terms, and posing $x_i = e^{V_i}$, we can write

$$P(i \mid \mathcal{C}) = \sum_m \frac{\left(\sum_j \alpha_{jm} e^{\mu_m V_j}\right)^{\frac{\mu}{\mu_m}}}{\sum_n \left(\sum_j \alpha_{jn} e^{\mu_n V_j}\right)^{\frac{\mu}{\mu_n}}} \frac{\alpha_{im} e^{\mu_m V_i}}{\sum_j \alpha_{jm} e^{\mu_m V_j}}, \tag{28}$$

which can nicely be interpreted as

$$P(i \mid \mathcal{C}) = \sum_m P(m \mid \mathcal{C}) P(i \mid m). \tag{29}$$

The issue of parameter identifiability is still to be addressed. In addition to the normalization of the Alternative Specific Constants (see Bierlaire, Lotan and Toint, 1997) and of the parameter μ, normalization of the α_{im} and μ_m parameters is also required. It is important to emphasize that constraints (10), (15), (18) and (20), proposed in the literature, are not necessary for the model validity (see Theorem 2). They are used to enable parameter identification, or to simplify the interpretation of the model. Abbé (2003) has shown that a proper normalization for the α parameters is

$$\sum_m \alpha_{jm}^{\frac{\mu}{\mu_m}} = e^{-\gamma} \quad \forall j \in \mathcal{C}, \tag{30}$$

where $\gamma \simeq 0.5772$ is Euler's constant.

4. Estimation procedure

The estimation procedures proposed by Small (1987) and Vovsha (1997) are based on heuristics. Small reduces the number of free parameters by imposing arbitrary restrictions on the parameters: $w_m = \frac{1}{M+1}, \forall m$, and $\rho_r = \rho, \forall r$. Vovsha proposes a complicated heuristic, where each observation is artificially substituted with n observations (Vovsha proposes $n = 100$).

Most of the time, the use of such heuristics is motivated by existing software packages, restricted to estimate simpler models. But the estimated parameters may be biased and sometimes even inconsistent with the theory. Instead, we prefer to exploit the closed form formula of the probability model (28) to perform a classical maximum likelihood estimation of the parameters.

Maximum likelihood estimation aims at identifying the set of parameters maximizing the probability that a given model perfectly reproduces the observations. It is a non-linear programming problem. The nature of the objective function and of the constraints determines the type of solution algorithm that must be used. The objective function of the maximum likelihood estimation problem for the Cross-Nested model is a non-linear function. In general, the function is not concave which significantly complicates the identification of a (global) maximum. Most non-linear programming algorithms (see Dennis and Schnabel, 1983; Bertsekas, 1999) are designed to identify local optima of the objective function. They require the availability of the derivatives of the objective function and of the constraints. As the Cross-Nested model has a closed-from, so does the log-likelihood function. Therefore, the analytical formula for the derivatives can be used. They are provided in the appendix.

There exists some meta-heuristics designed to identify global optima, like simulated annealing (Kirkpatrick, Gelatt and Vecchi, 1983; Rossier, Troyon and Liebling, 1986), tabu search (Glover, 1977; Hansen, 1986; Hansen and Jaumard, 1987) and variable neighborhood search (Hansen and Mladenovic, 1997). However, none of them can guarantee that the provided solution is indeed a global optimum. Therefore, whatever algorithm is preferred, starting it from different initial solutions is a good practice.

Constraints have to be imposed on parameters to be estimated. On the one hand, constraints defined by Theorem 2 guarantee the model validity. On the other hand, normalization constraints (such as (30)) are necessary for the model to be identifiable.

In the past, it was usually advised to explicitly incorporate normalization constraints (by setting a fixed value to some parameters), to ignore other constraints, and to use unconstrained optimization algorithms. The complexity of the CNL model, combined with the availability of efficient software packages for constrained optimization (Murtagh and Saunders, 1987; Conn, Gould and Toint, 1992; Lawrence, Zhou and Tits, 1997) now motivate the explicit management of constraints in the estimation process. Also, explicit constraints avoid meaningless values of the parameters to be generated during the iterations of the optimization algorithms.

A model estimation package called Biogeme (Bierlaire, 2003) has been developed, and is freely available from `biogeme.epfl.ch`. It is designed to estimate any model within the GEV model family. Non-linear utility functions can be handled. In particular, a specific scale parameter can be associated with different segments in the sample, and Box-Cox and Box-Tukey transforms can be applied to the attributes. Finally, any type of (continuous) constraint on the parameter can be defined. Biogeme proposes several optimization algorithms: CFSQP by Lawrence, Zhou and Tits (1997), DONLP2 by Spellucci (1993), SOLVOPT by Kuntsevich and Kappel (1997) and BIO, a recent implementation of Bierlaire (1995). A case study using Biogeme to estimate a CNL model in a mode choice SP/RP context is described by Bierlaire, Axhausen and Abbay (2001).

5. Conclusion and perspectives

The CNL model is appealing to capture complex situations where correlations cannot be handled by the Nested Logit model. Even with few alternatives and nests, the use of a CNL instead of a NL model may significantly improve the estimated model (Bierlaire, Axhausen and Abbay, 2001).

In this paper, a formal proof has been provided that CNL is indeed a member of the GEV family. Moreover, an estimation procedure based on classical non-linear programming techniques has been suggested to perform the log-likelihood estimation of the model, instead of the heuristics proposed in the literature. This procedure has been embedded in a new software package designed to estimate GEV models in general and CNL in particular. Also, derivatives of the log-likelihood function for the CNL model are provided.

We are currently conducting some research to adapt non-linear optimization procedures when the model is not identifiable (Bierlaire and Thémans, 2005), that is when no appropriate normalization has been performed. Indeed, as models become more sophisticated, the theoretical analysis of their identifiability is more and more difficult, and a numerical solution is desirable.

Appendix: A Derivatives

We provide here the derivatives of the log-likelihood function for GEV models in general, and for the Cross-Nested Logit model in particular. These formula's have been implemented in the Biogeme software package, and their validity has been checked against numerical finite difference approximations of the derivatives.

Given a sample of observations, the log-likelihood of the sample is

$$\mathcal{L} = \sum_{n \in \text{sample}} \ln P(i_n \mid C_n), \tag{31}$$

where i_n is the alternative actually chosen by individual n, C_n is the choice set, and

$$\ln P(i_n \mid C_n) = V_{i_n} + \ln G_{i_n} - \ln\left(\sum_{j \in \mathcal{C}_n} e^{V_j} G_j\right) \tag{32}$$

where $G_i = \partial G / \partial x_i$. In the following, we drop index n for the sake of simplification. All sums on j and k are over all alternatives in $\mathcal{C}_n$.

For any GEV model, if β_k is an unknown parameter to be estimated, we have

$$\frac{\partial \ln P(i \mid C)}{\partial \beta_k} = \frac{\partial V_i}{\partial \beta_k} + \frac{1}{G_i}\frac{\partial G_i}{\partial \beta_k} - \frac{\sum_j e^{V_j}\left(G_j \frac{\partial V_j}{\partial \beta_k} + \frac{\partial G_j}{\partial \beta_k}\right)}{\sum_k e^{V_k + \ln G_k}}, \tag{33}$$

with

$$\frac{\partial G_i}{\partial \beta_k} = \sum_j \frac{\partial G_i}{\partial x_j}\frac{\partial x_j}{\partial \beta_k},$$

where $\partial G_i / \partial x_j$ for the CNL model is given by (40) and (41). As $x_j = e^{V_j}$, we have

$$\frac{\partial G_i}{\partial \beta_k} = \sum_j e^{V_j}\frac{\partial G_i}{\partial x_j}\frac{\partial V_j}{\partial \beta_k}. \tag{34}$$

Note that this formula is sufficiently general to capture non-linear utility functions, such that $\partial V_j / \partial \beta_k$ is not necessarily constant. If λ_k is a model parameter, such as the μ_m or α_{jm} in (16), the derivative simplifies as

$$\frac{\partial \ln P(i \mid C)}{\partial \lambda_k} = \frac{1}{G_i}\frac{\partial G_i}{\partial \lambda_k} - \frac{\sum_j e^{V_j}\frac{\partial G_j}{\partial \lambda_k}}{\sum_k e^{V_k + \ln G_k}}, \tag{35}$$

where $\partial G_j / \partial \lambda_k$ for the CNL model is given by (42)–(45).

In the specific case of Cross-Nested logit model, we provide the first and second derivatives of (16) with respect to every parameter. The sums with index m are over all nests. The first derivative with respect to a variable x_i is given by

$$G_i = \frac{\partial G}{\partial x_i} = \mu \sum_m \alpha_{im} x_i^{\mu_m - 1} \left(\sum_j \alpha_{jm} x_j^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1}. \tag{36}$$

The first derivative with respect to the μ parameter is

$$\frac{\partial G}{\partial \mu} = \sum_m \frac{1}{\mu_m} y_m^{\frac{\mu}{\mu_m}} \ln(y_m), \tag{37}$$

where y_m is defined by (22). The first derivative with respect to the nest parameter μ_m is

$$\frac{\partial G}{\partial \mu_m} = \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 1} \left(\sum_j \alpha_{jm} x_j^{\mu_m} \ln(x_j) \right) - \frac{\mu}{\mu_m^2} y_m^{\frac{\mu}{\mu_m}} \ln(y_m) \tag{38}$$

and with respect to the α_{im} parameter is

$$\frac{\partial G}{\partial \alpha_{im}} = \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 1} x_i^{\mu_k}. \tag{39}$$

We now provide the second derivative with respect to x_i and x_j. If $i = j$, we have

$$\frac{\partial^2 G}{\partial x_i^2} = \frac{\partial G_i}{\partial x_i} = \sum_m \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 2} \alpha_{im} \mu_m x_i^{\mu_m - 2} \left(\left(\frac{\mu}{\mu_m} - 1 \right) \alpha_{im} \mu_m x_i^{\mu_m} + y_m(\mu_m - 1) \right) \tag{40}$$

and if $i \neq j$, we have

$$\frac{\partial^2 G}{\partial x_i \partial x_j} = \frac{\partial G_i}{\partial x_j} = \sum_m \mu_m \mu \left(\frac{\mu}{\mu_m} - 1 \right) \alpha_{im} \alpha_{jm} y_m^{\frac{\mu}{\mu_m} - 2} x_i^{\mu_m - 1} x_j^{\mu_m - 1} \tag{41}$$

where y_m is defined by (22).

The second derivative with respect to x_i and μ is

$$\frac{\partial^2 G}{\partial x_i \partial \mu} = \frac{\partial G_i}{\partial \mu} = \sum_m y_m^{\frac{\mu}{\mu_m} - 1} \alpha_{im} x_i^{\mu_m - 1} \left(1 + \frac{\mu}{\mu_m} \ln(y_m) \right). \tag{42}$$

The second derivative with respect to x_i and μ_m is

$$\begin{aligned} \frac{\partial^2 G}{\partial x_i \partial \mu_m} &= \frac{\partial G_i}{\partial \mu_m} \\ &= -\frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 1} \alpha_{im} x_i^{\mu_m - 1} - \frac{\mu^2}{\mu_m^2} y_m^{\frac{\mu}{\mu_m} - 1} \ln(y_m) \alpha_{im} x_i^{\mu_m - 1} \\ &\quad + \frac{\mu}{\mu_m} y_m^{\frac{\mu}{\mu_m} - 1} \alpha_{im} x_i^{\mu_m - 1} + \mu y_m^{\frac{\mu}{\mu_m} - 1} \alpha_{im} x_i^{\mu_m - 1} \ln(x_i). \end{aligned} \tag{43}$$

The second derivative with respect to x_i and α_{ik} is

$$\frac{\partial^2 G}{\partial x_i \partial \alpha_{ik}} = \frac{\partial G_i}{\partial \alpha_{ik}} = \mu x_i^{\mu_k - 1} y_k^{\frac{\mu}{\mu_k} - 1} \left(1 + \alpha_{ik} \left(\frac{\mu}{\mu_k} - 1 \right) y_k^{-1} x_i^{\mu_k} \right) \tag{44}$$

and with respect to x_i and α_{jk} $(i \neq j)$ is

$$\frac{\partial^2 G}{\partial x_i \partial \alpha_{jk}} = \frac{\partial G_i}{\partial \alpha_{jk}} = \mu \alpha_{ik} x_i^{\mu_k - 1} \left(\frac{\mu}{\mu_k} - 1 \right) y_k^{\frac{\mu}{\mu_k} - 2} x_j^{\mu_k} \tag{45}$$

where y_k is defined by (22).

Acknowledgments The author would like to thank Nicolas Antille for his help in the theoretical analysis, and Mats Weber, Michaël Thémans and two anonymous reviewers for their comments on an earlier version of the paper.

References

Abbé, E. (2003). "Behavioral Modeling: Generalized Extreme Value for Choice Models." Diploma Thesis. Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Abbé, E., M. Bierlaire, and T. Toledo. (2005). "Normalization and Correlation of Generalized Extreme Value Models." *Technical Report*, Ecole Polytechnique Fédérale de Lausanne (EPFL).

Ben-Akiva, M. and M. Bierlaire. (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. and M. Bierlaire. (2003). "Discrete Choice Models with Applications to Departure Time and Route Choice." In R. Hall (ed.), *Handbook of Transportation Science*, 2nd edn., Kluwer, pp. 7–37.

Ben-Akiva, M., M. Bierlaire, D. Burton, H.N. Koutsopoulos, and R. Mishalani. (2001). "Network State Estimation and Prediction for Real-Time Transportation Management Applications." *Networks and Spatial Economics* 1(3/4), 293–318.

Bertsekas, D.P. (1999). *Nonlinear Programming*, 2nd edn., Athena Scientific, Belmont.

Bhat, C. (1995). "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research Part B* 29, 471–483.

Bierlaire, M. (1995). "A Robust Algorithm for the Simultaneous Estimation of Hierarchical Logit Models." *GRT Report 95/3*, Department of Mathematics, FUNDP.

Bierlaire, M. (2003). "BIOGEME: A Free Package for the Estimation of Discrete Choice Models." *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M., K. Axhausen, and G. Abbay. (2001). "Acceptance of Model Innovation: The Case of the Swissmetro." *Proceedings of the 1st Swiss Transport Research Conference*.

Bierlaire, M., T. Lotan, and P.L. Toint. (1997). "On the Overspecification of Multinomial and Nested Logit Models Due to Alternative Specific Constants." *Transportation Science*, 31(4), 363–371.

Bierlaire, M., R. Mishalani, and M. Ben-Akiva. (2000). "General Framework for Dynamic Demand Simulation." *Technical Report RO-000223*, ROSO-DMA-EPFL Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne. http://roso.epfl.ch/mbi/demand-report.pdf.

Bierlaire, M. and M. Thémans. (2005). "Algorithmic Developments for the Estimate of Advanced Discrete Choice Models." *Proceedings of the 5th Swiss Transportation Research Conference*. www.strc.ch.

Chatterjee, K., N. Hounsell, P. Firmin, and P. Bonsall. (2002). "Driver Response to Variable Message Sign Information in London." *Transportation Research Part C,* 10, 149–169.

Conn, A.R., N.I.M. Gould, and P.L. Toint. (1992). "LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)." Number 17 in *Springer Series in Computational Mathematics*, Springer Verlag, Heidelberg, Berlin, New York.

Dennis, J.E. and R.B. Schnabel. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, USA.

Forinash, C.V. and F.S. Koppelman. (1993). "Application and Interpretation of Nested Logit Models of Intercity Mode Choice." *Transportation Research Record Issue*, 1413, 98–106.

Glover, F. (1977). "Heuristic for Integer Programming Using Surrogate Constraints." *Decision Sciences*, 8, 156–166.

Hansen, P. (1986). "The Steepest Ascent Mildest Descent Heuristic for Combinatorial Programming." *Technical Report*, Congress on Numerical Methods in Combinatorial Optimization. Capri, Italy.

Hansen, P. and B. Jaumard. (1987). "Algorithms for the Maximum Satisfiability Problem." *Technical Report* Rutgers University.

Hansen, P. and N. Mladenovic. (1997). "An Introduction to Variable Neighborhood Search." *Les Cahiers du GERAD*.

Kirkpatrick, S., C.D.J. Gelatt, and M.P. Vecchi. (1983). "Optimization by Simulated Annealing." *Science*, 220, 671–680.

Kuntsevich, A.V. and F. Kappel. (1997). *The Solver for Local Nonlinear Optimization Problems*. Institute for Mathematics, KarlFranzens University of Graz, Heinrichstr. 36, A-8010 Graz (Austria). www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/content.html.

Lawrence, C., J. Zhou, and A. Tits. (1997). "User's Guide for CFSQP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints." *Technical Report TR-94-16r1*, Institute for Systems Research, University of Maryland, College Park, MD 20742, 1997.

McFadden, D. (1978). "Modelling the Choice of Residential Location." In A. Karlquist et al. (eds.), *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, pp. 75–96.

Murtagh, B.A. and M.A. Saunders. (1987). MINOS 5.1 USER'S GUIDE, *Technical Report SOL83-20R*, Department of Operations Research, Stanford University, Stanford, USA.

Papola, A. (2004). "Some Developments on the Cross-Nested Logit Model." *Transportation Research B*, 38(9), 833–851.

Prashker, J. and S. Bekhor. (1999). "Stochastic User-Equilibrium Formulations for Extended-Logit Assignment Models." *Transportation Research Record*, 1676, 145–152.

Rossier, Y., M. Troyon, and T.M. Liebling. (1986). "Probabilistic Exchange Algorithms and Euclidean Traveling Salesman Problems." *OR Spektrum*, 8(3), 151–164.

Small, K. (1987). "A Discrete Choice Model for Ordered Alternatives." *Econometrica*, 55(2), 409–424.

Spellucci, P. (1993). *DONLP2 Users Guide*, Dept. of Mathematics, Technical University at Darmstadt, 64289 Darmstadt, Germany.

Swait, J. (2001). "Choice Set Generation Within the Generalized Extreme Value Family of Discrete Choice Models." *Transportation Research Part 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 Record*, 1607, 6–15.

Vovsha, P. and S. Bekhor. (1998). "The Link-Nested Logit Model of Route Choice: Overcoming the Route Overlapping Problem." *Transportation Research Record*, 1645, 133–142.

Wen, C.-H. and F.S. Koppelman. (2001). "The Generalized Nested Logit Model." *Transportation Research Part B*, 35(7), 627–641.

Williams, H. (1977). "On the Formation of Travel Demand Models and Economic Measures of User Benefit." *Environment and Planning*, 9A, 285–344.

Zeng, L. C. (2000). "A Heteroscedastic Generalized Extreme Value Discrete Choice Model." *Sociological Methods and Research*, 29, 118–144.