Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Part B 41 (2007) 795–808
TRANSPORTATION RESEARCH PART B
www.elsevier.com/locate/trb

Normalization and correlation of cross-nested logit models

E. Abbe [a,*], M. Bierlaire [b], T. Toledo [c]

[a] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 01239, United States
[b] Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[c] Transport Research Institute, Technion, Haifa 32000, Israel

Received 8 September 2005; received in revised form 17 September 2006; accepted 20 November 2006

Abstract

We address two fundamental issues associated with the use of cross-nested logit models. On the one hand, we justify the adequate normalization of the model proposed by Wen and Koppelman [Wen, C.-H., Koppelman, F.S., 2001. The generalized nested logit model. Transportation Research Part B 35(7), 627–641]. On the other hand, we provide an analysis of the correlation structure of the CNL, based on random utility theory. We evaluate the validity of the approximation proposed by Papola [Papola, A., 2004. Some developments on the cross-nested logit model. Transportation Research B 38(9), 833–851], both using theoretical arguments and numerical examples based on the exact formula of the correlation.

Keywords: Behavior models; Logit models; Generalized extreme-value; Correlations; Route choice; Model estimation

1. Introduction

The importance of demand analysis in transportation studies is increasingly critical. Discrete choice models provide a useful framework to capture the behavior of the actors within transportation systems and, consequently, to forecast travel demands. Recently, the cross-nested logit (CNL) model has received significant attention in the literature. Its structure is appealing since it can capture a wide range of correlation structures, while maintaining a closed form probability formula. The CNL model is therefore increasingly used in travel behavior applications such as mode choice (Vovsha, 1997; Bierlaire et al., 2001), departure time choice (Small, 1987) and route choice (Vovsha and Bekhor, 1998).

The CNL model is complicated, primarily for the following reasons. First, several formulations (and names) have been proposed in the literature, with associated normalization conditions. It is not always clear which one to adopt. Second, the variance–covariance matrix of the CNL is not simple to compute. Third, it is difficult to estimate its parameters in practice due to the presence of local maxima in the log-likelihood function.

As shown by Bierlaire (2006), various instances of the cross-nested logit model have been proposed in the literature. These formulations are generally equivalent, with some being more specific as they constrain some

* Corresponding author. Tel.: +1 617 253 4575.
E-mail addresses: eabbe@mit.edu (E. Abbe), michel.bierlaire@epfl.ch (M. Bierlaire), toledo@technion.ac.il (T. Toledo).

0191-2615/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trb.2006.11.006

parameters to fixed values. The formulations by Ben-Akiva and Bierlaire (1999), Wen and Koppelman (2001) and Papola (2004) are the most general. In this paper, we prefer a formulation which combines the GEV form of Ben-Akiva and Bierlaire (1999), and the simple normalization condition of Wen and Koppelman (2001).

Thanks to its closed form, the CNL may appear to be easy to estimate. Indeed, classical nonlinear programming methods based on derivatives, like Sequential Quadratic Programming (see e.g. Spellucci, 1998), are appropriate. The package BIOGEME (Bierlaire, 2003, Bierlaire, 2005) implements this method. Unfortunately, nonlinear programming methods converge towards *local* maxima of the log-likelihood function. They offer no guarantee to identify the global maximum. In practice, we observe a significant influence of the initial values provided to the algorithm on the estimated parameters.

In various applications, such as route choice analysis for instance, it is desirable to derive a CNL model which reproduces a given variance–covariance structure. Vovsha and Bekhor (1998) proposed the link-nested logit model, where the physical overlap of paths is used to define a CNL model. Papola (2004) generalizes this idea, and proposes a method to specify a CNL model reproducing any given homoscedastic covariance matrix. The procedure solves a system of equations, based on a conjecture about the covariance matrix of the CNL model. He also shows that a direct expression can be found without solving the system of equations, assuming that the covariance matrix is proportional to the utility function. Although this is only an approximation, this is the first paper in the literature providing operational formulas for the correlation of CNL.

In this paper we address two fundamental issues associated with the use of CNL models: the proper normalization of this model and its correlation structure (the issue of local maxima is left for future research). The rest of this paper is organized as follows. We present the CNL formulations in Section 2. In Section 3, we formally justify the validity of the normalization proposed by Wen and Koppelman (2001). The correlation structure of the CNL is analyzed in Section 4, where it is derived from the theory on generalized extreme value (GEV) models. We finally describe how the formulation can be used to derive CNL models from given variance–covariance matrices. Finally, we apply it on some illustrative examples, and compare the results with Papola's approach. Some technicalities about solving the system of equations defining the correlation, and about the route choice examples are detailed in Appendices A and B.

2. GEV models

The generalized extreme value (GEV) model was derived from the random utility model by McFadden et al. (1978). This general model consists of a large family of models that includes the multinomial logit (MNL), the nested logit (NL), the cross-nested logit (CNL) and the generalized nested logit (GNL) models. In GEV models, the probability that a given choice maker chooses alternative i within the choice set C is

$$P(i|C) = \frac{y_i G_i(y_1, \ldots, y_J)}{\mu G(y_1, \ldots, y_J)} = \frac{e^{V_i + \ln G_i(\ldots)}}{\sum_{j \in \mathscr{C}} e^{V_j + \ln G_j(\ldots)}}, \tag{1}$$

where $G_i = \partial G / \partial y_i$, J is the number of available alternatives, $y_i = e^{V_i}$, V_i is the systematic part of the utility function associated with alternative i, and G is a non-negative differentiable function defined on $\mathbb{R}^J_+$ which verifies some specific properties (see McFadden et al. (1978), Ben-Akiva and François (1983) and Ben-Akiva and Bierlaire (2003) for details). In his original paper, McFadden et al. (1978) defines the joint distribution of the random utility functions within a GEV model. The utility functions are modeled by a random vector of variables U defined by

$$U = V + \varepsilon, \tag{2}$$

where $V \in \mathbb{R}^{\mathbb{J}}$ and ε is a random vector of J variables with a cumulative distribution function (CDF) given by

$$F_{\varepsilon_1, \ldots, \varepsilon_J}(y_1, \ldots, y_J) = \exp(-G(e^{-y_1}, \ldots, e^{-y_J})). \tag{3}$$

It is well known that the multinomial logit and the nested logit models are instances of this model family, with

$$G(y_1, \ldots, y_J) = \sum_{j \in \mathscr{C}} y_j^{\mu} \tag{4}$$

for the multinomial logit and

$$G(y_1,\ldots,y_J)=\sum_{m=1}^{M}\left(\sum_{j=1}^{J_m}y_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}} \tag{5}$$

for the nested logit model with M nests containing J_m alternatives each.

Various formulations for the cross-nested logit model have been proposed in the literature. In this paper, we call a cross-nested logit model a GEV model based on the following generating function:

$$G(y_1,\ldots,y_J)=\sum_{m=1}^{M}\left(\sum_{j\in\mathscr{C}}(\alpha_{jm}^{1/\mu}y_j)^{\mu_m}\right)^{\frac{\mu}{\mu_m}}, \tag{6}$$

where

$$\alpha_{jm}\geqslant 0,\ \forall j,m,\quad \sum_{m=1}^{M}\alpha_{jm}>0,\ \forall j,\quad \mu>0,\quad \mu_m>0,\ \forall m,\quad \mu\leqslant\mu_m,\ \forall m. \tag{7}$$

This formulation leads to the following probability model, using $y_i=\mathrm{e}^{V_i}$:

$$P(i|\mathscr{C})=\sum_{m=1}^{M}\frac{\left(\sum_{j\in\mathscr{C}}\alpha_{jm}^{\mu_m/\mu}\mathrm{e}^{\mu_m V_j}\right)^{\frac{\mu}{\mu_m}}}{\sum_{n=1}^{M}\left(\sum_{j\in\mathscr{C}}\alpha_{jn}^{\mu_n/\mu}\mathrm{e}^{\mu_n V_j}\right)^{\frac{\mu}{\mu_n}}}\frac{\alpha_{im}^{\mu_m/\mu}\mathrm{e}^{\mu_m V_i}}{\sum_{j\in\mathscr{C}}\alpha_{jm}^{\mu_m/\mu}\mathrm{e}^{\mu_m V_j}}. \tag{8}$$

For each j and m, the parameter α_{jm} is interpreted as the level of membership of alternative j to nest m. The nested logit model is a special case, where $\alpha_{jm}=1$ if alternative j belongs to nest m, and 0 otherwise.

The name *cross-nested* seems to be due to Vovsha (1997), who applies this model to mode choice in Tel Aviv. Vovsha's model is similar to the Ordered GEV model proposed by Small (1987). Ben-Akiva and Bierlaire (1999) and Wen and Koppelman (2001) propose more general formulations, the latter being called the "Generalized nested logit" model.

The flexible correlation structure of the CNL model is useful in a wide range of applications. For example, it has 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. Prashker and Bekhor (1999) discuss the use of route choice models based on a simplified CNL model within the stochastic user equilibrium traffic assignment context. In another application, Swait (2001) suggests a 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 also be derived (see, for instance, Bhat, 1995; Zeng, 2000 and Koppelman and Sethi, 2005).

The CNL model is appealing for its ability to capture a wide variety of correlation structures. Bierlaire (2006) compares various formulations, and identifies the sufficient conditions for the CNL to be a GEV model. Papola (2004) has conjectured that a specific CNL model can be obtained for any given homoscedastic variance–covariance matrix.

3. Normalization

The CNL model requires a normalization of the underlying structural parameters. In this section we formally derive this normalization.

Consider a cross-nested logit based on the utility functions (2) where ε is a random vector with a cumulative distribution function (CDF) given by (3) where G is defined by (6). The marginal distribution of ε_j, $j\in\mathscr{C}$ is given by

$$F_{\varepsilon_j}(y_j)=\exp\left[-\exp\left\{-\mu\left(y_j-\frac{\ln\left(\sum_{m=1}^{M}\alpha_{jm}\right)}{\mu}\right)\right\}\right]. \tag{9}$$

Therefore, the marginal distribution of (3) for the cross-nested model has an extreme value (EV) distribution with location parameter

$$\frac{\ln\left(\sum_{m=1}^{M}\alpha_{jm}\right)}{\mu}$$

and scale parameter μ. Therefore,

$$E[\varepsilon_j] = \frac{\ln\left(\sum_{m=1}^{M}\alpha_{jm}\right) + \gamma}{\mu}, \tag{10}$$

where $\gamma \approx 0.5772$ is the Euler constant. It is important to emphasize that the expected value of ε_j depends on the α_{jm} parameters and, consequently, may vary from one alternative to the next.

It is critical to define a normalization constraint on the α parameters in order to have unbiased models satisfying, for all $j \in \mathscr{C}$,

$$E[U_j] = V_j + E[\varepsilon_j] = V_j + K,$$

where K is a constant independent from alternative j. Note, that in the special case of the multinomial logit and the nested logit models, we have

$$E[\varepsilon_j] = \frac{\gamma}{\mu}, \tag{11}$$

which is always constant across alternatives.

If the $E[\varepsilon_j]$ are not constant across alternatives the model would still be unbiased provided that a full set of alternative specific constants (ASC) were included in the model (that is, $J-1$ ASCs). But the interpretation of these constants would not be compatible with other models calibrated on the same data. Also, standard corrections of the constants, accounting for selection bias in the sample (see e.g. Bierlaire et al., 2006), would not apply as such. Furthermore, some models do not contain a full set of constants. This is typically the case when the number of alternatives is large (like in residential location choice, or route choice analysis), or when data comes from unlabeled stated preferences experiments. In this case, the model would be artificially biased, leading to incorrect prediction of market shares. Thus, a proper normalization of the model parameters is important. Namely, if

$$\sum_{m=1}^{M}\alpha_{jm} = c, \quad \forall j \in \mathscr{C}, \tag{12}$$

where c is a constant which does not depend on j, then

$$E[U_j] = V_j + E[\varepsilon_j] = V_j + \frac{\ln c + \gamma}{\mu}, \quad \forall j \in \mathscr{C}.$$

Clearly, a value of $c = 1$ seems natural, as it yields to an expected value of γ/μ, similar to the MNL and NL models.

We note that, if the formulation proposed by Ben-Akiva and Bierlaire (1999) is preferred, that is the model based on

$$G(y_1, \ldots, y_J) = \sum_{m=1}^{M}\left(\sum_{j\in\mathscr{C}}\alpha_{jm}y_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}}, \tag{13}$$

the proper normalization is

$$\sum_{m=1}^{M}\alpha_{jm}^{\frac{\mu}{\mu_m}} = c, \quad j \in \mathscr{C}, \tag{14}$$

which must be preferred to the normalization originally proposed in their paper. This normalization is inconvenient, involving nonlinear constraints, even if μ is constrained to 1.

We finally note that the formulation proposed by Wen and Koppelman (2001) is equivalent to (6) with the additional assumption that μ has been normalized to 1. This assumption does not lead to any loss of generality.

4. Variance–covariance structure

The variance–covariance structures of MNL and NL are well-known. Because of the IIA property, the MNL has a diagonal variance–covariance matrix. For the nested logit models, only alternatives belonging to the same nest are correlated, and the correlation is defined by

$$\mathrm{Corr}(U_i, U_j) = \left(1 - \left(\frac{\mu}{\mu_m}\right)^2\right)\delta_m(i,j), \tag{15}$$

where

$$\delta_m(i,j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are both in nest } m, \\ 0 & \text{otherwise.} \end{cases}$$

The CNL model, which allows alternatives to belong to multiple nests, exhibits a more general correlation structure. We start by presenting an interpretation of the CNL structure in terms of underlying nested logit models, which provides some insight about the error structure of the CNL model.

Theorem 1. *Considering a cross-nested logit with a cumulative distribution function (CDF) given by (3) and G defined by (6). It is equivalent to the model defined by*

$$U_j = \max_{m=1,\dots,M} \widehat{U}_{jm}, \tag{16}$$

where

$$\widehat{U}_{jm} = V_j + \frac{\ln \alpha_{jm}}{\mu} + \varepsilon_{jm} \tag{17}$$

and for any m, the joint distribution of $(\varepsilon_{1m}, \dots, \varepsilon_{Jm})$ *is given by (18),*

$$F_{\varepsilon_{1m},\dots,\varepsilon_{Jm}}(y_1, \dots, y_J) = \exp\left(-\left(\sum_{j\in\mathscr{C}} \mathrm{e}^{-\mu_m y_j}\right)^{\frac{\mu}{\mu_m}}\right), \tag{18}$$

whereas ε_{im} *and* ε_{jn} *are independent (for any i and j), as long as* $m \neq n$.

Proof. Let $\widehat{\varepsilon}_{jm} = \frac{\ln \alpha_{jm}}{\mu} + \varepsilon_{jm}$ and $\varepsilon_j^* = \max_m \widehat{\varepsilon}_{jm}$. We show that ε^* is distributed as the ε of the CNL model, that is

$$F_{\varepsilon_1^*,\dots,\varepsilon_J^*}(y_1, \dots, y_J) = F_{\varepsilon_1,\dots,\varepsilon_J}(y_1, \dots, y_J). \tag{19}$$

The CDF of ε^* is

$$\begin{aligned} F_{\varepsilon_1^*,\dots,\varepsilon_J^*}(y_1, \dots, y_J) &= \Pr\left(\varepsilon_1^* \leqslant y_1, \dots, \varepsilon_J^* \leqslant y_J\right) = \Pr\left(\max_m \widehat{\varepsilon}_{1m} \leqslant y_1, \dots, \max_m \widehat{\varepsilon}_{Jm} \leqslant y_J\right) \\ &= \Pr(\widehat{\varepsilon}_{11} \leqslant y_1, \dots, \widehat{\varepsilon}_{1M} \leqslant y_1, \dots, \widehat{\varepsilon}_{J1} \leqslant y_J, \dots, \widehat{\varepsilon}_{Jm} \leqslant y_J) \\ &= \prod_{m=1}^{M} \Pr\left(\varepsilon_{1m} \leqslant y_1 - \frac{\ln \alpha_{1m}}{\mu}, \dots, \varepsilon_{Jm} \leqslant y_J - \frac{\ln \alpha_{Jm}}{\mu}\right) \\ &= \prod_{m=1}^{M} F_{\varepsilon_{1m},\dots,\varepsilon_{Jm}}\left(y_1 - \frac{\ln \alpha_{1m}}{\mu}, \dots, y_J - \frac{\ln \alpha_{Jm}}{\mu}\right). \end{aligned}$$

Therefore,

$$\ln F_{\varepsilon_1^*,\ldots,\varepsilon_J^*}(y_1,\ldots,y_J) = \sum_{m=1}^{M} \ln F_{\varepsilon_{1m},\ldots,\varepsilon_{Jm}}\left(y_1 - \frac{\ln \alpha_{1m}}{\mu},\ldots,y_J - \frac{\ln \alpha_{Jm}}{\mu}\right) = -\sum_{m=1}^{M}\left(\sum_{j\in\mathscr{C}} \mathrm{e}^{-\mu_m\left(y_j - \frac{\ln \alpha_{jm}}{\mu}\right)}\right)^{\frac{\mu}{\mu_m}}$$

$$= -\sum_{m=1}^{M}\left(\sum_{j\in\mathscr{C}} (\alpha_{jm}^{\frac{1}{\mu}}\mathrm{e}^{-y_j})^{\mu_m}\right)^{\frac{\mu}{\mu_m}} = -G(\mathrm{e}^{-y_1},\ldots,\mathrm{e}^{y_J}),$$

which proves (19). □

Distribution (18) corresponds to a GEV model with

$$G(y_1,\ldots,y_J) = \left(\sum_{j\in\mathscr{C}} y^{\mu_m}\right)^{\frac{\mu}{\mu_m}},$$

which is the mth term of the G function of the nested logit model (5). As a direct consequence of Theorem 1, we can state the following result.

Corollary 2. *Under the same hypothesis as Theorem 1, we have*

$$\mathrm{Corr}(U_i,U_j) = \mathrm{Corr}\left(\max_m \widehat{\varepsilon}_{im}, \max_m \widehat{\varepsilon}_{jm}\right),$$

where

$$\mathrm{Corr}(\widehat{\varepsilon}_{im},\widehat{\varepsilon}_{jn}) = \left(1 - \left(\frac{\mu}{\mu_m}\right)^2\right)\delta_{m,n}.$$

Equivalently,

$$\mathrm{Corr}(U_i,U_j) = \mathrm{Corr}\left(\max_m\left(\frac{\ln \alpha_{im}}{\mu} + \varepsilon_{im}\right), \max_m\left(\frac{\ln \alpha_{jm}}{\mu} + \varepsilon_{jm}\right)\right),$$

where

$$\mathrm{Corr}(\varepsilon_{im},\varepsilon_{jn}) = \left(1 - \left(\frac{\mu}{\mu_m}\right)^2\right)\delta_{m,n}.$$

So, the relation between the overall CNL correlation and the underlying NL correlations is made via a *maximum* operator. This is actually consistent with the GEV framework.

The CNL correlation itself is obtained from the joint CDF of the utilities, that is

$$\mathrm{Corr}(U_i,U_j) = \frac{6\mu^2}{\pi^2}\int\int_{\mathbb{R}^2} x_i x_j \partial^2_{x_i x_j} F_{\varepsilon_i,\varepsilon_j}(x_i,x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j - \frac{6\gamma^2}{\pi^2}, \tag{20}$$

where

$$F_{\varepsilon_i,\varepsilon_j}(x_i,x_j) = \exp\left(-\sum_{m=1}^{M}\left((\alpha_{im}^{1/\mu}\mathrm{e}^{-x_i})^{\mu_m} + (\alpha_{jm}^{1/\mu}\mathrm{e}^{-x_j})^{\mu_m}\right)^{\frac{\mu}{\mu_m}}\right). \tag{21}$$

The CNL model includes a large number of structural parameters. In many cases, it may be useful to calculate values of these parameters prior to estimating the other model parameters in order to reduce the dimensionality of the estimation problem. For example, in a route choice application Vovsha and Bekhor (1998) used the network topology to calculate α_{im} and μ_m. The exact correlation structure of CNL derived above can be used in such calculations. For a given correlation structure, we can derive the associated CNL model by computing the structural parameters reproducing the given correlation. It is done by solving a system of equations defined by (20) and (12).

For a set of J alternatives, the system includes $J(J-1)/2$ equations for the correlations, and J equations for the normalization conditions, that is $J(J-1)/2+J$ equations, and so this is the maximum number of parameters that can be identified. If the CNL model contains M nests, and each alternative belongs to all nests, there are JM α-parameters and M μ-parameters to be estimated. Consequently, all parameters will be uniquely specified from the correlation structure only when J is even and $M=J/2$.

When all the parameters cannot be identified, additional equations must be added to overcome the underdetermination of the system. The simplest of such equations would consist in assigning arbitrary values to unidentified parameters. However, it is crucial to verify that the extended system of equations has a solution which is consistent with the theory (that is, verify conditions (7)). In the general case, it can only be done empirically after the solution has been obtained. Also, we emphasize that the scale of the model may vary depending on the arbitrary choice made. Indeed, the systems of equations impose values for the correlation, not for the covariance. However, this issue is important only when several models are compared together.

The variance–covariance structure of the CNL was also analyzed by Papola (2004), who proposes the following approximation for a CNL model derived from (6):

$$\widehat{\mathrm{Corr}}(U_i,U_j)=\sum_{m=1}^{M}\alpha_{im}^{1/2}\alpha_{jm}^{1/2}\left(1-\left(\frac{\mu}{\mu_m}\right)^2\right). \tag{22}$$

Papola (2004) validates this formula using limit cases, when the CNL model collapses to a NL model.

Actually, this conjecture simplifies the approach by replacing the maximum operator in Corollary 2 by a *linear* interpolation of the underlying NL correlation (22) with regard to the nests m. The weights are chosen to get for $i=j$

$$\sum_{m=1}^{M}\alpha_{im}^{1/2}\alpha_{im}^{1/2}=\sum_{m=1}^{M}\alpha_{im}=1$$

from (12), which should represent the variance of U_i. The approximation errors due to this simplification may sometimes be important. This is analyzed in Section 5.

5. Illustrative examples

In this section we illustrate the behavior of the CNL model in terms of the correlations among alternatives and the choice probabilities using three simple examples. In all cases we compare the results to those obtained using Papola's approximation.

5.1. First example

We first consider the simple CNL structure shown in Fig. 1.

We are interested in the correlation between alternatives 1 and 2, both belonging to nest A. We assume in this example, and later examples, that the scale parameters are equal, $\mu_A=\mu_B=\mu_m$. The root node has scale parameter μ. Note that application of the normalization conditions (12) for this structure results in $\alpha_{A1}=\alpha_{B3}=1$ and $\alpha_{B2}=1-\alpha_{A2}$. Fig. 2 shows values of the correlation and the one computed using Papola's approximation as a function of α_{A2} for selected values of μ/μ_m. The correlations were computed from (20) via numerical integration and (22), respectively.

This example illustrates that Papola's approximation overestimates the correlation between the two alternatives. The overestimation, which can be significant, increases when the value of μ/μ_m decreases (i.e. the nesting structure is more and more significant in explaining the choice between the alternatives). For a given value of μ/μ_m, the error in Papola's approximation is rather small for low values of α_{A2} and increases with the value of α_{A2}. It attains its maximum value for α_{A2} in the range of 0.3–0.4 (in this example) and then decreases again as α_{A2} continues to increase. We note that Papola and Marzano (2005) also report such an overestimation. Papola's approximation was constructed such that it correctly captures the correlations in the special cases of multinomial logit and nested logit, and so in this example it is exact when $\alpha_{A2}=0$ or $\alpha_{A2}=1$.

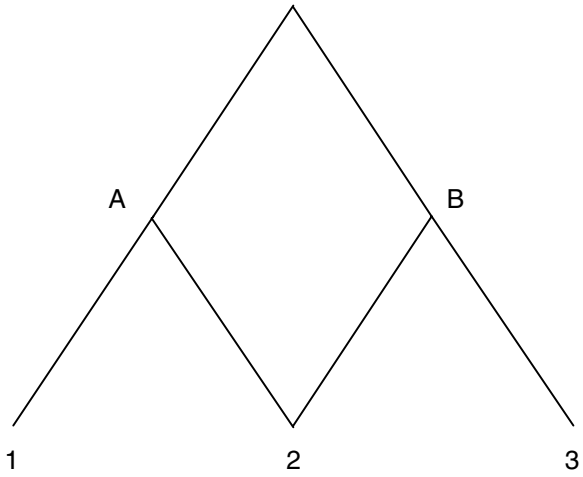

Fig. 1. Simple cross-nested structure.

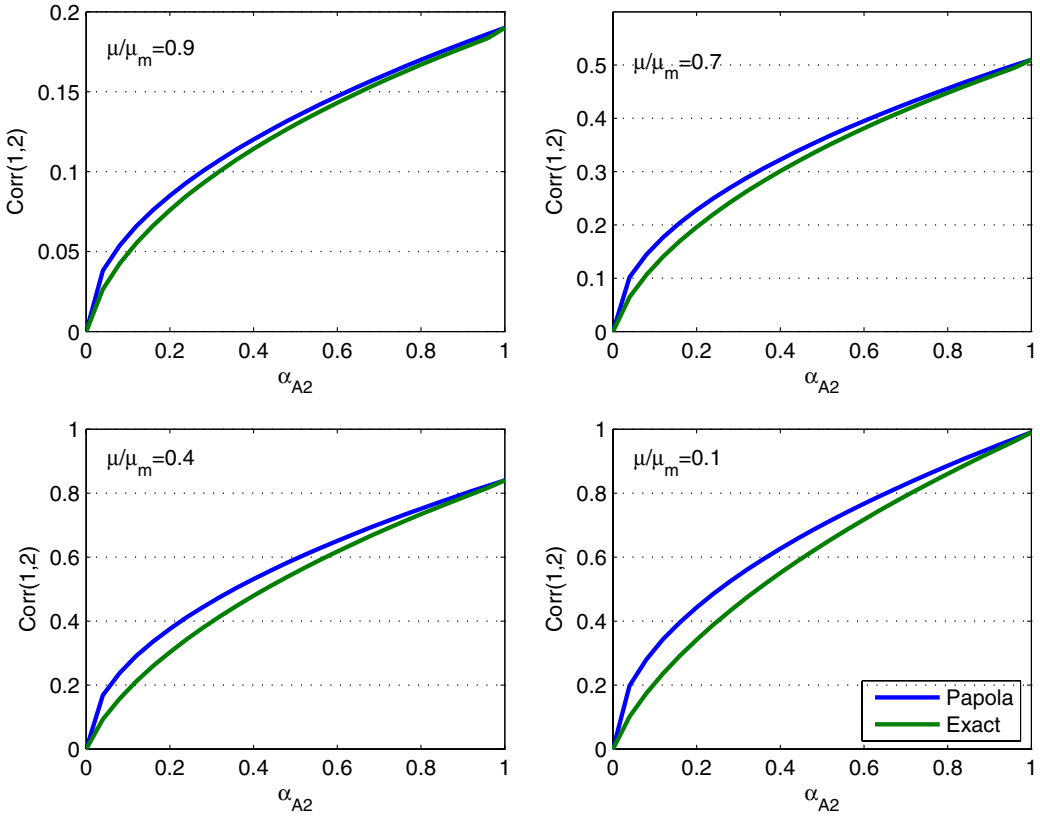

Fig. 2. Comparison of correlation with Papola's approximation.

5.2. Second example

In the next example, we consider the CNL structure shown in Fig. 3. Compared to the previous example, this structure has an additional link from nest B to alternative 1, and so alternatives 1 and 2, now share two common nests. For this structure, $\alpha_{B3} = 1$, $\alpha_{B1} = 1 - \alpha_{A1}$ and $\alpha_{B2} = 1 - \alpha_{A2}$. Fig. 4 shows correlation values

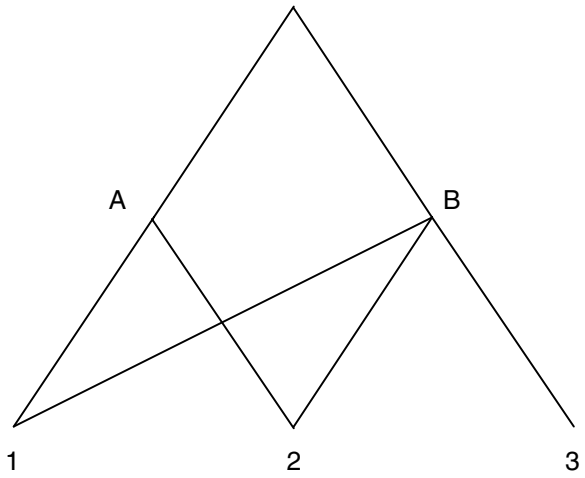

Fig. 3. Second cross-nested structure.

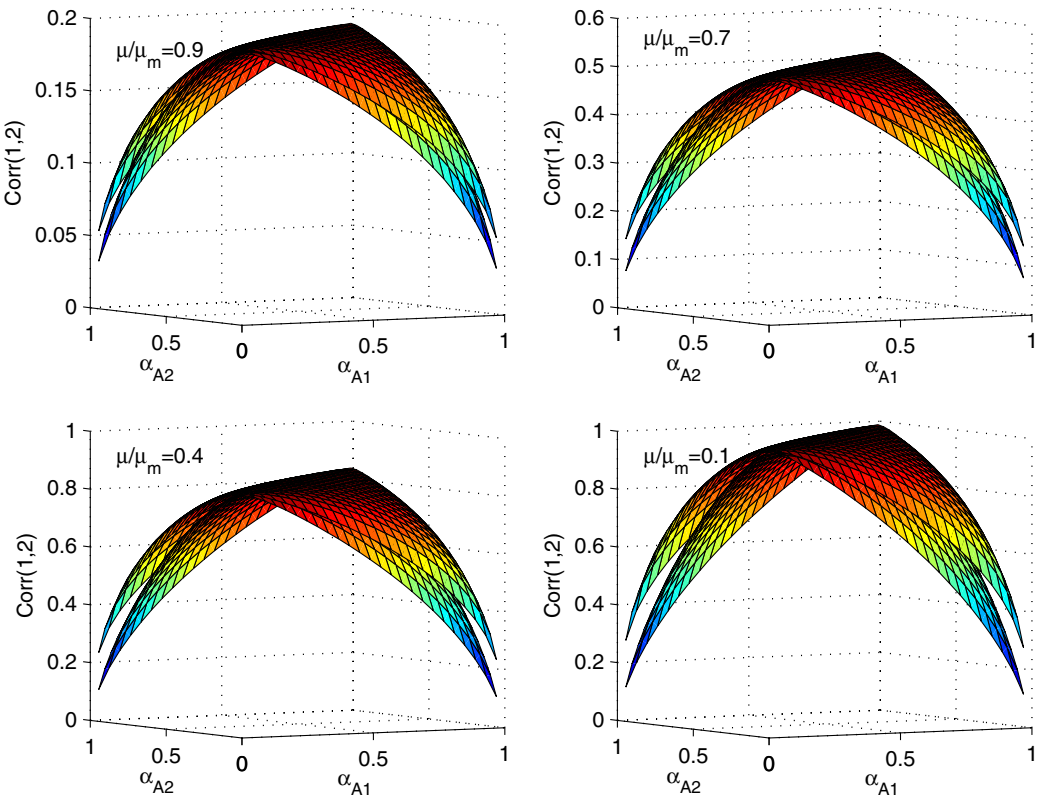

Fig. 4. Second CNL: comparison of correlation and Papola's approximation.

and Papola's approximation as a function of α_{A1} and α_{A2} for selected values of μ/μ_m. In all cases the graph on top represents Papola's expression. Thus, as with the previous example, Papola's expression overestimates the correlation between the two alternatives. Furthermore, depending on the values of the α's, the difference may be larger than in the case that the two alternatives share only one nest.

This example allows us to also examine the impact of the allocation of the two alternatives to nests on the correlation. Fig. 5 shows the correlation between alternatives 1 and 2, and Papola's approximation as a

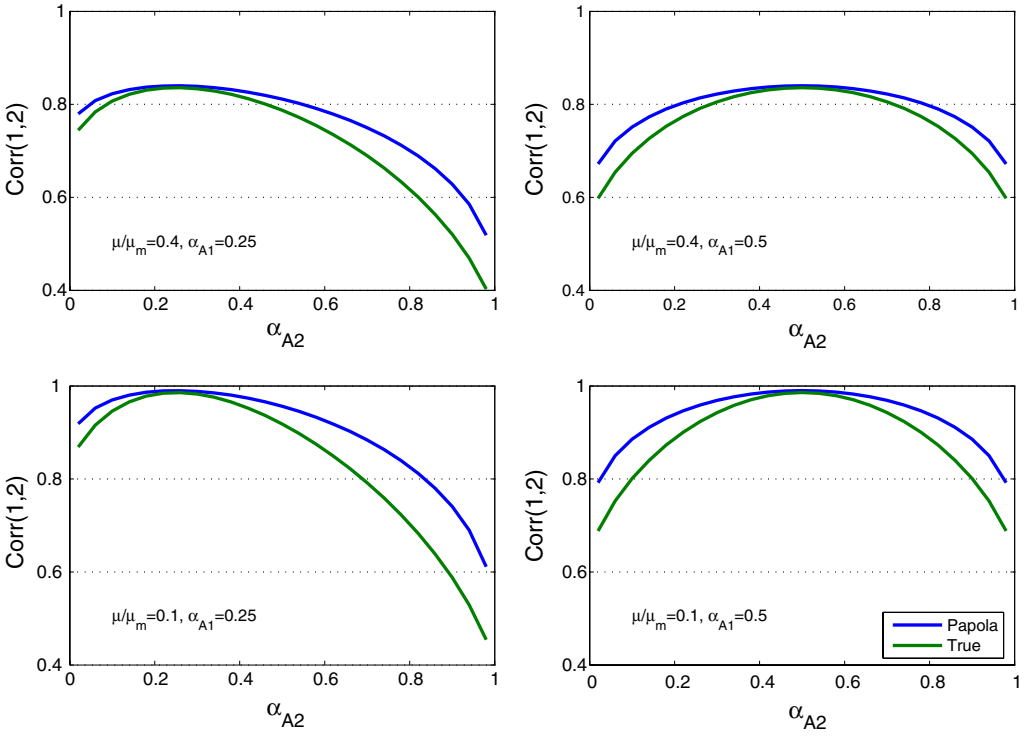

Fig. 5. Layers from Fig. 4.

function of α_{A2} for given values of μ/μ_m and α_{A1}. The figure illustrates that the correlation between the alternatives is highest when the two have identical allocations to each nest, and generally reduces when the allocations increasingly differ. While Papola's approximation captures this general trend it increasingly overestimates the correlation, when the allocations differ.

Note also that Papola's approximation is exact when $\alpha_{A1} = \alpha_{A2}$. Actually, this happens in the specific case where all the nest parameters are the same and, for each nest m, all α_{im} parameters are also the same, that is $\mu_m = \mu_0$, for all nests m, and $\alpha_{im} = \alpha_m$, for each alternative i, where μ_0 is independent of m and α_m is independent of i. In this case, the joint CDF of the utilities of the two alternatives given by (21) can be simplified:

$$F_{\varepsilon_i,\varepsilon_j}(y_i,y_j) = \exp\left(-\sum_m \left(\alpha_m^{\frac{\mu_0}{\mu}} e^{-x_i\mu_0} + \alpha_m^{\frac{\mu_0}{\mu}} e^{-x_j\mu_0}\right)^{\frac{\mu}{\mu_0}}\right) = \exp\left(-\sum_m \left(\alpha_m^{\frac{\mu_0}{\mu}} (e^{-x_i\mu_0} + e^{-x_j\mu_0})\right)^{\frac{\mu}{\mu_0}}\right)$$
$$= \exp\left(-\sum_m \alpha_m (e^{-x_i\mu_0} + e^{-x_j\mu_0})^{\frac{\mu}{\mu_0}}\right) = \exp\left(-(e^{-x_i\mu_0} + e^{-x_j\mu_0})^{\frac{\mu}{\mu_0}}\right).$$

This expression is identical to the joint CDF of the utilities of two alternatives in the same nest in a NL structure. Thus, the correlation between the two is given by

$$\mathrm{Corr}(U_i,U_j) = 1 - \left(\frac{\mu}{\mu_0}\right)^2.$$

Papola's approximation for this case is given by

$$\widehat{\mathrm{Corr}}(U_i,U_j) = \sum_m \alpha_{im}^{\frac{1}{2}} \alpha_{jm}^{\frac{1}{2}} \left(1 - \left(\frac{\mu}{\mu_m}\right)^2\right) = \sum_m \alpha_m \left(1 - \left(\frac{\mu}{\mu_0}\right)^2\right) = 1 - \left(\frac{\mu}{\mu_0}\right)^2$$

and is exact.

5.3. A route choice example

The previous two examples illustrate the behavior of the correlation as a function of the values of the nesting structure parameters. They also show that the overestimation of correlations is inherent to Papola's approximation. However, in most practical applications we are interested in the choice probabilities of the various alternatives and not in the correlations among these alternatives per se. The next example demonstrates the calculation of nesting parameters, for a given correlation structure, based on the correlation expression and shows the difference in the prediction of choice probabilities using the exact expression and Papola's approximation. We consider a route choice problem for the network shown in Fig. 6. The mean travel times on each link are indicated in the figure. There are three routes from the origin node to the destination node (denoted as O and D, respectively, in the figure) in this network: {A, D}, {A, C, E} and {B, E}. We denote these as routes 1, 2 and 3 respectively. The mean travel times on each route are equal to 1 unit. We assume that route choices are based solely on these travel times, and that the standard deviation of travel times is proportional to the mean in each link (i.e. $\mathrm{var}(tt_i) = \sigma^2 * \bar{tt}_i$). We further assume that travel times on the various links are independent of each other (i.e. $\mathrm{cov}(tt_i, tt_j) = 0$). Under these assumptions, the correlation matrix of the travel times on the three routes is given by

$$\begin{pmatrix} 1 & & \\ a & 1 & \\ 0 & b & 1 \end{pmatrix}.$$

We use a link-route CNL structure (Vovsha and Bekhor, 1998) to model route choice in this network. The structure shown in Fig. 7 has the links (A–E) at the upper level and the three routes at the lower level. Each route is connected to all the links it consists of. Thus, the model structure has seven α's that need to

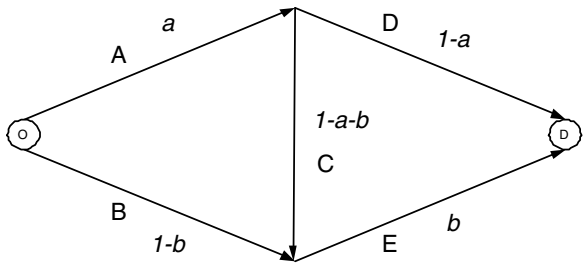

Fig. 6. Simple network for route choice example.

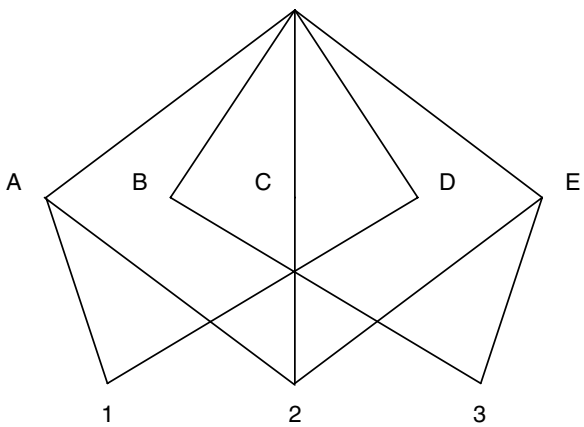

Fig. 7. Cross-nested structure of the route choice example.

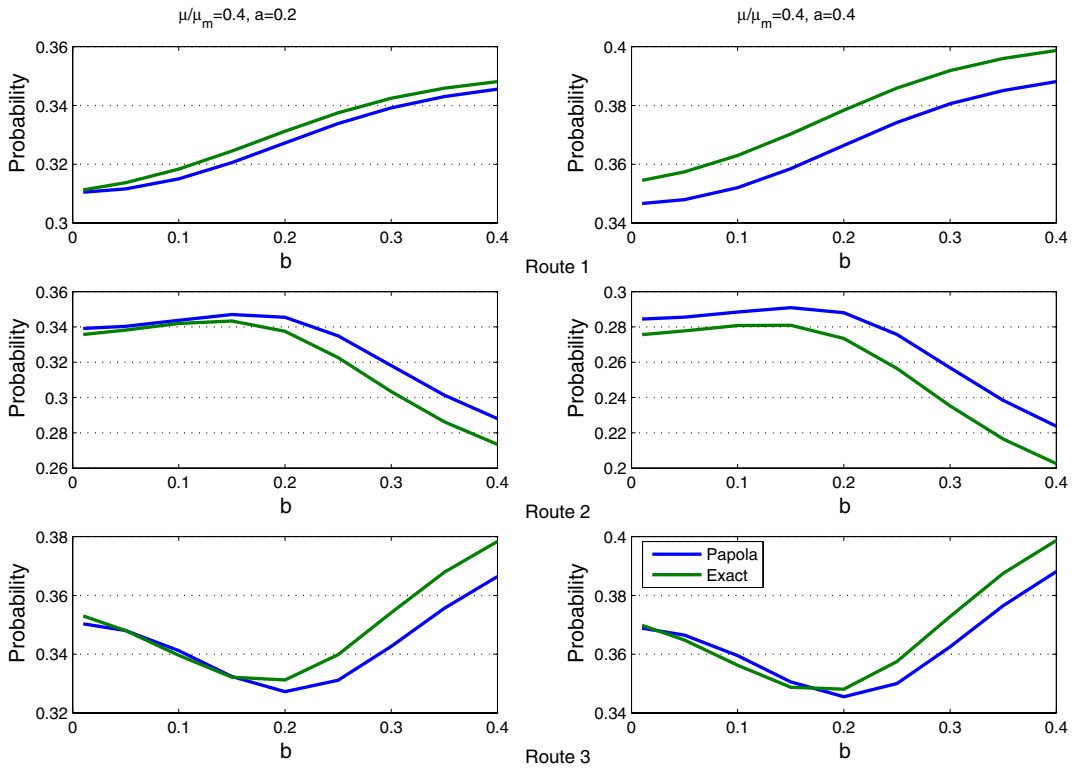

Fig. 8. Comparison of choice probability between Papola's approximation and exact formulation.

be estimated in addition to the ratio μ/μ_m. We would like to use the assumed correlation structure and the normalization conditions to estimate these parameters. However, these only provide five equations in this case, and so we arbitrarily set $\alpha_{A2} = \alpha_{C2} = \alpha_{E2} = 1/3$ and $\mu/\mu_m = 0.4$. We now calculate the remaining four parameters using the four correlation and normalization equations (the normalization for alternative 2 has been used; see Appendices A and B for details).

Reproduction of a correlation structure is not sufficient to specify the model fully because the error variance and parameters of the systematic utilities have not yet been set. In a practical application the structural parameters values (α's and μ's) that were obtained by the solution of the above system of equations will now be fixed and used in estimating the values of the remaining parameters. Here we skip this step and use these parameter values to calculate choice probabilities (Recall that the systematic utilities are equal for all three alternatives). Fig. 8 shows the predicted choice probabilities for the three routes in this example for different values of a and b. With the values set above, no parameters values could produce correlations higher than 0.4. The results show that the probability values predicted using Papola's approximation differ by up to 2%. Furthermore, the difference between the two predictions generally increases with the level of correlation between the alternatives.

6. Conclusion and perspectives

In this paper, we have addressed two important issues related to the cross-nested logit model: normalization and correlation structure. Exploiting the GEV theory, we have emphasized that, contrarily to MNL and NL models, the expected value of the error terms in a CNL are not necessarily equal, and we have presented and proved a proper normalization, which appears to be a slight generalization of Wen and Koppelman (2001). We have then derived the formula for the correlation between two alternatives in a CNL model.

On some illustrative examples, we have compared that formulation with Papola's approximation. The latter ends up being a fairly good approximation for CNL models having only bipolar shared alternatives. For such

models, the use of the exact expression via the integral expression is possible too. When having two nests sharing the alternatives, the approximation stays reasonably good. It is not clear though how good it can stay when considering CNL models with several nests, as it is often the case in practical applications, such as in route choice problems. In general, it seems that Papola's approximation overestimate the correlation, and may bias the choice probability provided by the model. However, any such bias does not seem to be large in the present examples.

Consequently, in order to derive a CNL model with a given correlation structure, we recommend to use the exact formulation derived in this paper. This requires finding the solution of a system of equations involving numerical integration. This can be achieved using mathematical tools such as Matlab (Moled, 2004), or numerical procedures described in Press et al. (2002).

Acknowledgments

We would like to thank Andrea Papola and Vittorio Marzano for interesting discussions on these issues. We are also thankful to four anonymous reviewers who provided us with useful comments about a previous version of this paper.

Appendix A. Implementation notes

While the solution of the system of equations defined by (12) and (20) only requires standard techniques, it is not necessarily trivial. We provide in this appendix some practical comments that may be useful in implementing the results discussed above.

- The evaluation of (20) must be done numerically. The most popular numerical integration algorithm is based on the adaptive Simpson quadrature algorithm. Its implementation is described in details by Press et al. (2002, Section 4.2) and is used by the *quad* function of Matlab 7. Note that a double integral is required, which can be directly performed in Matlab 7 using the function *dblquad*.
- In theory, the bounds of integration in (20) are $-\infty$ and $+\infty$. In practice, numerical procedures require finite values. If the values are too high (in magnitude), this may cause numerical problems, and will anyway increase the computing time. If the values are too low, the tail of the distribution will be missed, and the value will be biased. For our tests, we have systematically used -10 and 10.
- The best procedure for solving the system of nonlinear equations is probably Broyden's method, described in detail by Dennis and Schnabel (1996) and Press et al. (2002, Section 9.7). It does not seem to be implemented in Matlab 7 as such. The procedure `fsolve` in Matlab 7 is (by default) a trust-region algorithm using a finite differences approximation of the Jacobian. If the number of equations is high, the algorithm proposed by Bierlaire and Crittin, 2006 is appropriate.

Appendix B. Some details on the route choice examples

In this appendix we provide the complete system of equations and the additional assumptions made in order to obtain a determined system for the route choice example in Section 5.3. The CNL structure for this problem is shown in Fig. 7. The structural parameters are α_{A1}, α_{D1}, α_{A2}, α_{C2}, α_{E2}, α_{B3}, α_{E3}, μ and μ_m. To set the scale of the model we set $\mu = 1$ and so only solve for the ratio μ/μ_m.

The correlation equations for this problem are

$$\mathrm{Corr}(U_1, U_2) = a, \tag{23}$$

$$\mathrm{Corr}(U_2, U_3) = b \tag{24}$$

with $\mathrm{Corr}(\cdot)$ defined by (20). The normalization equations are given by

$$\alpha_{A1} + \alpha_{D1} = 1, \tag{25}$$

$$\alpha_{A2} + \alpha_{C2} + \alpha_{E2} = 1, \tag{26}$$

$$\alpha_{B3} + \alpha_{E3} = 1. \tag{27}$$

In order to eliminate the under-determination we arbitrarily made the following assumptions: $\alpha_{A2} = \alpha_{C2} = \alpha_{E2} = 1/3$ (which satisfy Eq. (26)) and $\mu/\mu_m = 0.4$. Thus the final system to be solved consists of Eqs. (23)–(25) and (27), with the unknowns α_{A1}, α_{D1}, α_{B3}, α_{E3}.

References

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

Ben-Akiva, M., Bierlaire, M., 2003. Discrete choice models with applications to departure time and route choice. In: Hall, R. (Ed.), Handbook of Transportation Science, second ed. Kluwer Academic Publishers, Boston/Dordrecht/London, pp. 7–37.

Ben-Akiva, M., François, B., 1983. μ homogeneous generalized extreme value model, Working paper, Department of Civil Engineering, MIT, Cambridge, MA.

Bhat, C., 1995. A heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part B 29, 471–483.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland. www.strc.ch.

Bierlaire, M., 2005. An introduction to BIOGEME version 1.4. biogeme.epfl.ch.

Bierlaire, M., 2006. A theoretical analysis of the cross-nested logit model. Annals of Operations Research 144 (1), 287–300.

Bierlaire, M., Crittin, F., 2006. Solving noisy large scale fixed point problems and systems of nonlinear equations. Transportation Science 40 (1), 44–63.

Bierlaire, M., Axhausen, K., Abay, G., 2001. Acceptance of modal innovation: the case of the Swissmetro. In: Proceedings of the 1st Swiss Transportation Research Conference, Ascona, Switzerland. www.strc.ch.

Bierlaire, M., Bolduc, D., McFadden, D., 2006. The estimation of generalized extreme value models from choice-based samples, Technical report TRANSP-OR 060810, Transport and Mobility Laboratory, ENAC, EPFL, Lausanne, Switzerland.

Dennis, J.E., Schnabel, R.B., 1996. Numerical methods for unconstrained optimization and nonlinear equations, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Koppelman, F., Sethi, V., 2005. Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior. Transportation Research Part B 39 (9), 825–853.

McFadden, D., 1978. Modelling the choice of residential location. In: Karlquist, A. et al. (Eds.), Spatial Interaction Theory and Residential Location. North-Holland, Amsterdam, pp. 75–96.

Moled, C.B., 2004. Numerical Computing with Matlab. Society for Industrial and Applied Mathematics (SIAM).

Papola, A., 2004. Some developments on the cross-nested logit model. Transportation Research B 38 (9), 833–851.

Papola, A., Marzano, V., 2005. A specification procedure of the CNL model reproducing any homoskedastic covariance matrix. In: Proceedings of the European Transport Conference, Strasbourg, France.

Prashker, J., Bekhor, S., 1999. Stochastic user-equilibrium formulations for extended-logit assignment models. Transportation Research Record 1676, 145–152.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2002. Numerical Recipes in C++: The Art of Scientific Computing, second ed. Cambridge University Press.

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

Spellucci, P., 1998. An SQP method for general nonlinear programs using only equality constrained subproblems. Mathematical Programming 82, 413–448.

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., 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., Koppelman, F.S., 2001. The generalized nested logit model. Transportation Research Part B 35 (7), 627–641.

Zeng, L.C., 2000. A heteroscedastic generalized extreme value discrete choice model. Sociological Methods and Research 29, 118–144.