

# Normalization and correlation of cross-nested logit models

E. Abbe\*      M. Bierlaire<sup>†</sup>      T. Toledo<sup>‡</sup>

August 15, 2005

## 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 (2001). On the other hand, we provide an analysis of the correlation structure of the CNL, based on random utility theory. We show that the conjecture by Papola (2004) about the correlation structure of the CNL is incorrect, both using theoretical arguments and numerical examples based on the exact formula of the correlation.

## 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,

---

\*Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Email: eabbe@mit.edu

<sup>†</sup>Ecole Polytechnique Fédérale de Lausanne, Inst. of Mathematics, CH-1015 Lausanne, Switzerland. Email: michel.bierlaire@epfl.ch

<sup>‡</sup>Transport Research Institute, Technion - Israel Institute of Technology

while maintaining a closed form probability formula. The CNL model is therefore increasingly used in travel behavior applications such as mode choice (Vovsha, 1997 and Bierlaire et al., 2001), departure time choice (Small, 1987) and route choice (Vovsha and Bekhor, 1998).

The CNL model is a complicated model, namely 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, it is difficult to estimate its parameters in practice due to the presence of local maxima in the log-likelihood function and non trivial normalization constraints. Third, the variance-covariance matrix of the CNL is not simple to compute.

As shown by Bierlaire (forthcoming), 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 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 et al., 2004) implements this method. Unfortunately, nonlinear programming methods are converging towards *local* maxima of the loglikelihood 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, like 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 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, when the covariance matrix is proportional to the utility function.

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 we show that the conjecture proposed by Papola (2004) is incorrect, and we derive the correct formulation from the theory on Generalized Extreme Value (GEV) models. We finally use the formulation to derive CNL models from given variance-covariance matrices.

## 2 GEV models

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 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:

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

where  $G_i = \frac{\partial G}{\partial y_i}$ ,  $J$  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, 1978 or Ben-Akiva and Bierlaire, 2003 for details).

In his original paper, McFadden (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

$$U = V + \varepsilon \quad (2)$$

where  $V \in \mathbb{R}^J$  and  $\varepsilon$  is a random vector of  $J$  variables with a Cumulative Distribution Function (CDF) given by

$$F_{\varepsilon_1, \dots, \varepsilon_J}(y_1, \dots, y_J) = e^{-G(e^{-y_1}, \dots, e^{-y_J})}. \quad (3)$$

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

$$G(y) = \sum_{j \in \mathcal{C}} y_j^\mu \quad (4)$$

for the multinomial logit and

$$G(y) = \sum_{m=1}^M \left( \sum_{j=1}^{J_m} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \quad (5)$$

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

**Definition 1** *The cross-nested logit model is a GEV model based on the following generating function*

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

where  $\alpha_{jm} \geq 0$ ,  $\forall j, m$ ,  $\sum_{m=1}^M \alpha_{jm} > 0$ ,  $\forall j$ ,  $\mu > 0$ ,  $\mu_m > 0$ ,  $\forall m$ ,  $\mu \leq \mu_m$ ,  $\forall m$ .

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

$$P(i|\mathcal{C}) = \sum_{m=1}^M \frac{\left( \sum_{j \in \mathcal{C}} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_j} \right)^{\frac{\mu}{\mu_m}}}{\sum_{n=1}^M \left( \sum_{j \in \mathcal{C}} \alpha_{jn}^{\mu_n/\mu} e^{\mu_n V_j} \right)^{\frac{\mu}{\mu_n}}} \frac{\alpha_{im}^{\mu_m/\mu} e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_j}}. \quad (7)$$

For each  $j$  and  $m$ , the parameter  $\alpha_{jm}$  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 context. In another application, 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 and Zeng, 2000, Koppelman and Sethi, 2005).

This model is appealing for its ability to capture a wide variety of correlation structures. Bierlaire (forthcoming) 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. We show in Section 4 that this result does not hold in general.

### 3 Normalization

The CNL model requires a normalization of the underlying structural parameters. In this section we formally derive this normalization. We first analyze the marginal distribution of (3) for the cross-nested model, and we show that it has an extreme value (EV) distribution.

*Lemma 1 Consider a cross-nested logit based on the utility functions (2) where  $\varepsilon$  is a random vector with a Cumulative Distribution Function (CDF) given by (3) where  $G$  is defined by (6). The marginal*

distribution of  $\varepsilon_j$ ,  $j \in \mathcal{C}$  is given by

$$F_{\varepsilon_j}(\mathbf{y}_j) = \exp \left[ - \exp \left\{ -\mu \left( \mathbf{y}_j - \frac{\ln \left( \sum_{m=1}^M \alpha_{jm} \right)}{\mu} \right) \right\} \right]. \quad (8)$$

**Proof.** We have

$$\begin{aligned} \lim_{\mathbf{y}_k \rightarrow +\infty, k \neq j} G(e^{-\mathbf{y}_1}, \dots, e^{-\mathbf{y}_J}) &= \sum_{m=1}^M (\alpha_{jm}^{1/\mu} e^{-\mathbf{y}_j})^\mu \\ &= e^{-\mathbf{y}_j \mu} \sum_{m=1}^M \alpha_{jm} \\ &= \exp \left( -\mathbf{y}_j \mu + \ln \sum_{m=1}^M \alpha_{jm} \right). \end{aligned}$$

Since the marginal distribution is defined by

$$F_{\varepsilon_k}(\mathbf{y}_j) = \lim_{\mathbf{y}_k \rightarrow +\infty, k \neq j} F_{\varepsilon_1, \dots, \varepsilon_J}(\mathbf{y}_1, \dots, \mathbf{y}_J),$$

the result is obtained using (3) and the fact that the exponential function is continuous.  $\square$

The previous lemma shows that the marginal distribution of  $\varepsilon_j$  follows an EV distribution with location parameter

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

and scale parameter  $\mu$ . Therefore,

$$E[\varepsilon_j] = \frac{\ln \left( \sum_{m=1}^M \alpha_{jm} \right) + \gamma}{\mu}, \quad (9)$$

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

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

$$E[\mathbf{U}_j] = \mathbf{V}_j + E[\varepsilon_j] = \mathbf{V}_j + \mathbf{K},$$

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

$$E[\varepsilon_j] = \frac{\gamma}{\mu}, \quad (10)$$

which is always constant across alternatives.

The model would be unbiased if 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, would not apply as such. Also, some models do not contain a full set of constants. This is typically the case when the number of alternatives is very large (like in house location choice, or route choice analysis), or when data comes from unlabelled stated preferences experiments. In this case, the model would be artificially biased, leading to incorrect prediction of market shares.

**Corollary 2** *Consider a cross-nested logit based on the utility functions (2) where  $\varepsilon$  is a random vector with a Cumulative Distribution Function (CDF) given by (3) and  $G$  is defined by (6). If*

$$\sum_{m=1}^M \alpha_{jm} = c, \quad \forall j \in \mathcal{C}, \quad (11)$$

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

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

Clearly, a value of  $c = 1$  seems natural, as it yields to an expected value of  $\gamma/\mu$ , like for the MNL and NL models.

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

$$G(y_1, \dots, y_J) = \sum_{m=1}^M \left( \sum_{j \in \mathcal{C}} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \quad (12)$$

the proper normalization is

$$\sum_{m=1}^M \alpha_{jm}^{\frac{\mu}{\mu_m}} = c, \quad j \in \mathcal{C}, \quad (13)$$

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

We finally note that the formulation proposed by Wen and Koppelman (2001) is (6) where  $\mu$  has been normalized to 1, which can always be done without loss of generality.

## 4 Variance-covariance structure

The variance-covariance structure of MNL and NL is 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

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

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 provides a more general correlation structure. The variance-covariance structure of the CNL has been analyzed by Papola (2004), who proposes the following conjecture. For a CNL model (definition 1),

$$\text{Corr}(U_i, U_j) = \sum_{m=1}^M \alpha_{im}^{1/2} \alpha_{jm}^{1/2} \left(1 - \left(\frac{\mu}{\mu_m}\right)^2\right). \quad (14)$$

Papola (2004) validates this formula using limit cases, when the CNL model collapses to a NL model, that is for any  $m$ ,  $1 \leq m \leq M$ , for any  $i \neq j$ ,



$1 \leq i, j \leq J$ ,

$$\lim_{\substack{\alpha_{im} \rightarrow 1, \alpha_{jm} \rightarrow 1 \\ \alpha_{j1} \rightarrow 0, \alpha_{i1} \rightarrow 0, \forall l \neq m}} \text{Corr}(U_i, U_j) = \left(1 - \left(\frac{\mu}{\mu_m}\right)^2\right).$$

So the conjecture is a **linear** interpolation of the underlying NL correlation 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 (11), which should represent the variance of  $U_i$ . We now show that while it can in some cases provide a reasonable simple approximation, this conjecture is only an approximation. 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. This result helps understanding why the Papola's conjecture is not exact.

**Theorem 3** *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*

$$U_j = \max_{m=1, \dots, M} \hat{U}_{jm} \quad (15)$$

where

$$\hat{U}_{jm} = V_j + \frac{\ln \alpha_{jm}}{\mu} + \varepsilon_{jm} \quad (16)$$

and  $\varepsilon_{jm}$  are such that  $\varepsilon_{jm}$  is independent of  $\varepsilon_{ln}$ , for any  $m \neq n$  such that  $1 \leq m, n \leq M$ , for any  $j, l$  such that  $1 \leq j, l \leq J$ , and for any given  $m$  such  $1 \leq m \leq M$ ,  $\varepsilon_{jm}$  are distributed as

$$F_{\varepsilon_{1m}, \dots, \varepsilon_{jm}}(y_1, \dots, y_J) = e^{-(\sum_{j \in C} e^{-\mu_m y_j})^{\frac{\mu}{\mu_m}}}. \quad (17)$$

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

$$F_{\varepsilon_1^*, \dots, \varepsilon_J^*}(y_1, \dots, y_J) = F_{\varepsilon_1, \dots, \varepsilon_J}(y_1, \dots, y_J). \quad (18)$$

The CDF of  $\varepsilon^*$  is

$$\begin{aligned}
F_{\varepsilon_1^*, \dots, \varepsilon_J^*}(\mathbf{y}_1, \dots, \mathbf{y}_J) &= \Pr(\varepsilon_1^* \leq \mathbf{y}_1, \dots, \varepsilon_J^* \leq \mathbf{y}_J) \\
&= \Pr\left(\max_m \hat{\varepsilon}_{1m} \leq \mathbf{y}_1, \dots, \max_m \hat{\varepsilon}_{Jm} \leq \mathbf{y}_J\right) \\
&= \Pr(\hat{\varepsilon}_{11} \leq \mathbf{y}_1, \dots, \hat{\varepsilon}_{1M} \leq \mathbf{y}_1, \dots, \hat{\varepsilon}_{J1} \leq \mathbf{y}_J, \dots, \hat{\varepsilon}_{JM} \leq \mathbf{y}_J) \\
&= \prod_{m=1}^M \Pr\left(\varepsilon_{1m} \leq \mathbf{y}_1 - \frac{\ln \alpha_{1m}}{\mu}, \dots, \varepsilon_{Jm} \leq \mathbf{y}_J - \frac{\ln \alpha_{Jm}}{\mu}\right) \\
&= \prod_{m=1}^M F_{\varepsilon_{1m}, \dots, \varepsilon_{Jm}}\left(\mathbf{y}_1 - \frac{\ln \alpha_{1m}}{\mu}, \dots, \mathbf{y}_J - \frac{\ln \alpha_{Jm}}{\mu}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\ln F_{\varepsilon_1^*, \dots, \varepsilon_J^*}(\mathbf{y}_1, \dots, \mathbf{y}_J) &= \sum_{m=1}^M \ln F_{\varepsilon_{1m}, \dots, \varepsilon_{Jm}}\left(\mathbf{y}_1 - \frac{\ln \alpha_{1m}}{\mu}, \dots, \mathbf{y}_J - \frac{\ln \alpha_{Jm}}{\mu}\right) \\
&= -\sum_{m=1}^M \left(\sum_{j \in \mathcal{C}} e^{-\mu_m \left(\mathbf{y}_j - \frac{\ln \alpha_{jm}}{\mu}\right)}\right)^{\frac{\mu}{\mu_m}} \\
&= -\sum_{m=1}^M \left(\sum_{j \in \mathcal{C}} (\alpha_{jm}^{\frac{1}{\mu}} e^{-\mathbf{y}_j})^{\mu_m}\right)^{\frac{\mu}{\mu_m}} \\
&= -G(e^{-\mathbf{y}_1}, \dots, e^{-\mathbf{y}_J}),
\end{aligned}$$

which proves (18).

□

Distribution (17) corresponds to a GEV model with

$$G(\mathbf{y}_1, \dots, \mathbf{y}_J) = \left(\sum_{j \in \mathcal{C}} \mathbf{y}_j^{\mu_j}\right)^{\frac{\mu}{\mu_m}}$$

which is the  $m$ th term of the G function of the nested logit model (5).

As a direct consequence of theorem 3, we can state the following result.

**Corollary 4** *Under the same hypothesis as Theorem 3, we have*

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

where

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

Equivalently,

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

where

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

We now see that corollary 4 is actually not corroborating the conjecture (14). In fact, the proposed expression was claiming that the relation between underlying NL correlations and the overall CNL correlations is **linear**, whereas our corollary shows it is a relation made via a **maximum** operator.

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

$$\text{Corr}(U_i, U_j) = \frac{6\mu^2}{\pi^2} \iint_{\mathbb{R}^2} x_i x_j \partial_{x_i x_j}^2 F_{\varepsilon_i, \varepsilon_j}(x_i, x_j) dx_i dx_j - \frac{6\gamma^2}{\pi^2}, \quad (19)$$

where

$$F_{\varepsilon_i, \varepsilon_j}(x_i, x_j) = e^{-\sum_{m=1}^M \left( (\alpha_{im}^{1/\mu} e^{-x_i})^{\mu m} + (\alpha_{jm}^{1/\mu} e^{-x_j})^{\mu m} \right)^{\frac{\mu}{\mu_m}}}. \quad (20)$$

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  $\alpha_{im}$  and  $\mu_m$ . The exact correlation structure of CNL derived above can be used in such calculations. For a given correlation structure, the structural parameters can be computed by solving a system of equations defined by (19) and (11).

Note that 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 number of parameters that can be identified. If the CNL model contains  $M$  nests, there are  $JM$

$\alpha$ -parameters and  $M$   $\mu$ -parameters to be estimated. Consequently, all parameters will be uniquely specified from the correlation structure only when  $J$  is even and  $M = J/2$ .

## 4.1 Implementation notes

Solving the system of equations, although a standard technique, is not necessarily trivial. We conclude this section by providing some practical hints in order to help practitioners using better the results from the paper.

- The evaluation of (19) 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 (19) 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 (forthcoming) is appropriate.

## 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 obtained using the exact correlation to those obtained using Papola's expression.

### 5.1 First example

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

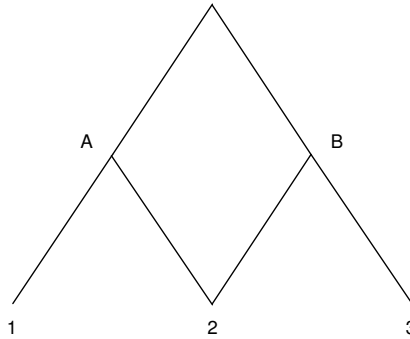


Figure 1: Simple cross-nested structure

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  $\mu$ . Note that application of the normalization conditions (11) for this structure results in  $\alpha_{A1} = \alpha_{B3} = 1$  and  $\alpha_{B2} = 1 - \alpha_{A2}$ . Figure 2 shows values of the exact correlation and the one computed using Papola's conjecture as a function of  $\alpha_{A2}$  for selected values of  $\mu/\mu_m$ . The correlations were computed from (19) via numerical integration and (14), respectively. This example illustrates that Papola's conjecture systematically overestimates the exact correlation between the two alternatives. The overestimation, which can be significant, increases when the value of  $\mu/\mu_m$  decreases (i.e. the nesting structure is more and more significant in explaining the

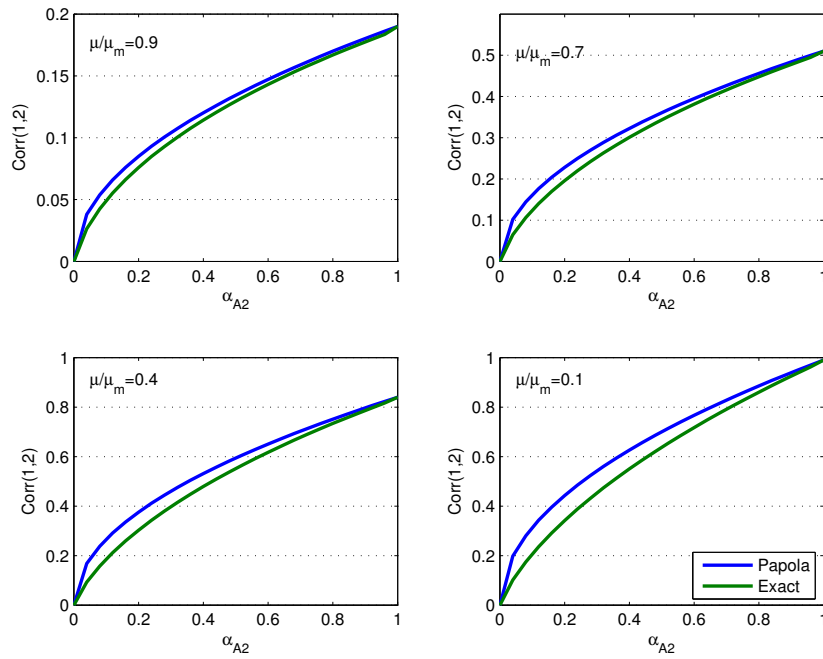


Figure 2: Comparison of Papola's conjecture and exact correlation

choice between the alternatives). For a given value of  $\mu/\mu_m$ , the error in Papola's conjecture is rather small for low values of  $\alpha_{A2}$  and increases with the value of  $\alpha_{A2}$ . It attains its maximum value for  $\alpha_{A2}$  in the range of 0.3-0.4 (in this example) and then decreases again as  $\alpha_{A2}$  continues to increase. Papola's conjecture 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  $\alpha_{A2} = 1$ .

## 5.2 Second example

In the next example, we consider the CNL structure shown in Figure 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  $\alpha_{B2} = 1 - \alpha_{A2}$ . Figure 4 shows correlation values using Papola's conjecture and the exact expression as a function of  $\alpha_{A1}$  and  $\alpha_{A2}$  for selected values of  $\mu/\mu_m$ . In all cases the graph on top represents Papola's expression. Thus, as with the previous example, Papola's expression systematically overestimates the correlation between alternatives 1 and 2. Furthermore, depending on the values of the  $\alpha$ 's, the difference may be larger than in the case that the two alternatives share only one nest.

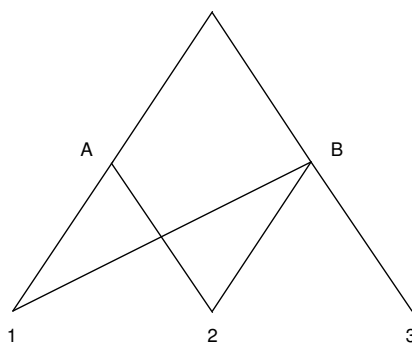


Figure 3: Second cross-nested structure

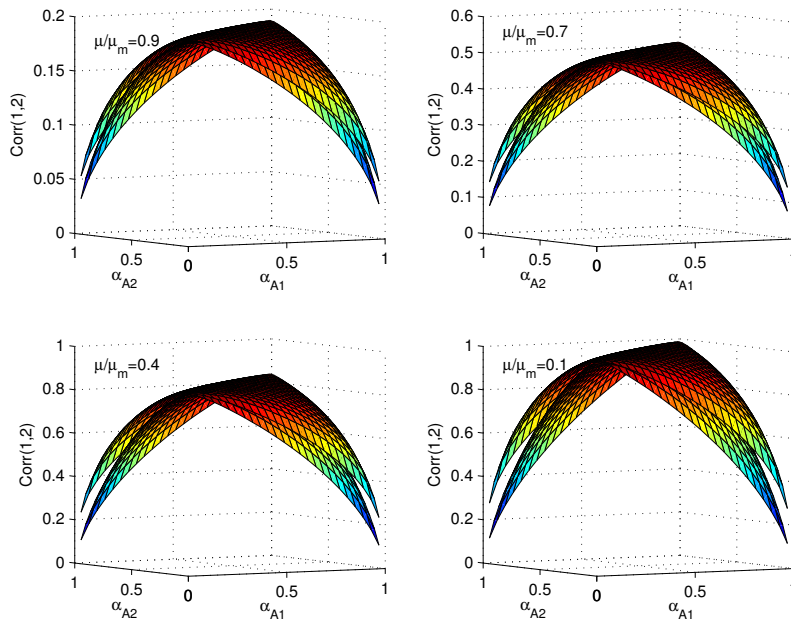


Figure 4: Second CNL: comparison of correlation between Papola's conjecture and exact formulation



This example allows us to also examine the impact of the allocation of the two alternatives to nests on the correlation. Figure 5 shows the correlation between alternatives 1 and 2, computed with the exact expression and Papola's conjecture as a function of  $\alpha_{A2}$  for given values of  $\mu/\mu_m$  and  $\alpha_{A1}$ . 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 conjecture captures this general trend it increasingly overestimates the exact correlation, when the allocations differ. Note also that Papola's conjecture is exact when  $\alpha_{A1} = \alpha_{A2}$ .

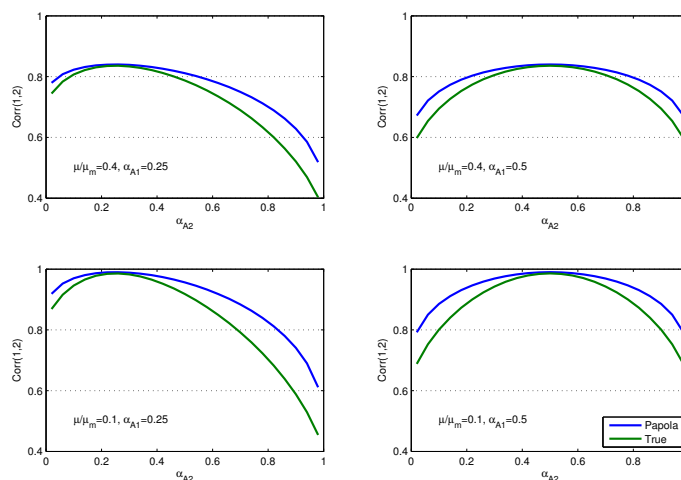


Figure 5: Layers from Figure 4

### 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 conjecture. However, in most practical applications we are interested in the choice prob-

abilities 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 conjecture. We consider a route choice problem for the network shown in Figure 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.  $\text{var}(tt_i) = \sigma^2 * \bar{tt}_i$ ). We further assume that travel times on the various links are independent of each other (i.e.  $\text{cov}(tt_i, tt_j) = 0$ ). Under these assumptions, the correlation matrix of the travel times on the three routes is given by:

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

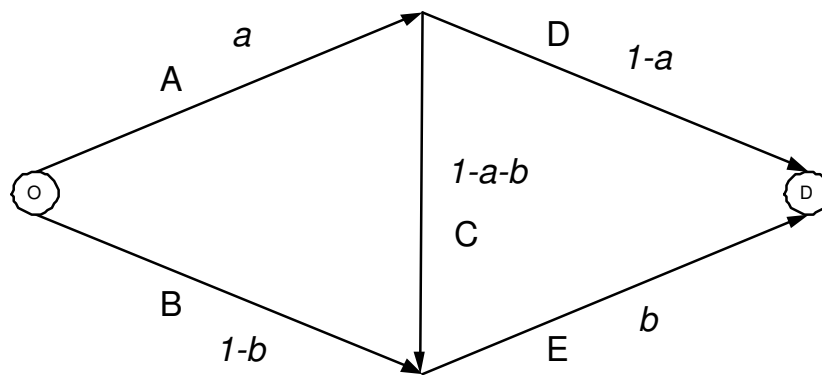


Figure 6: Simple network for route choice example

We use a link-route CNL structure (Vovsha and Bekhor, 1998) to model

route choice in this network. The structure shown in figure 7 has the links (A through 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  $\alpha$ 's that need to be estimated in addition to the ratio  $\mu/\mu_m$ . 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). These parameter values are then used to calculate choice probabilities (Recall that the systematic utilities are equal for all three alternatives). Figure 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 conjecture differ by up to 2% compared to those predicted using the exact expression. 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 a proper normalization, which appears to be a slight generalization of Wen and Koppelman (2001). We have then derived the correct formula for the correlation between two alternatives in a CNL model, and shown that the conjecture proposed by Papola (2004) is not correct. We have used numerical examples to show how Papola's approximation seems to systematically overestimate the correlation, and may bias the choice probability provided by the model, although such bias does not seem to be dramatic on our examples.

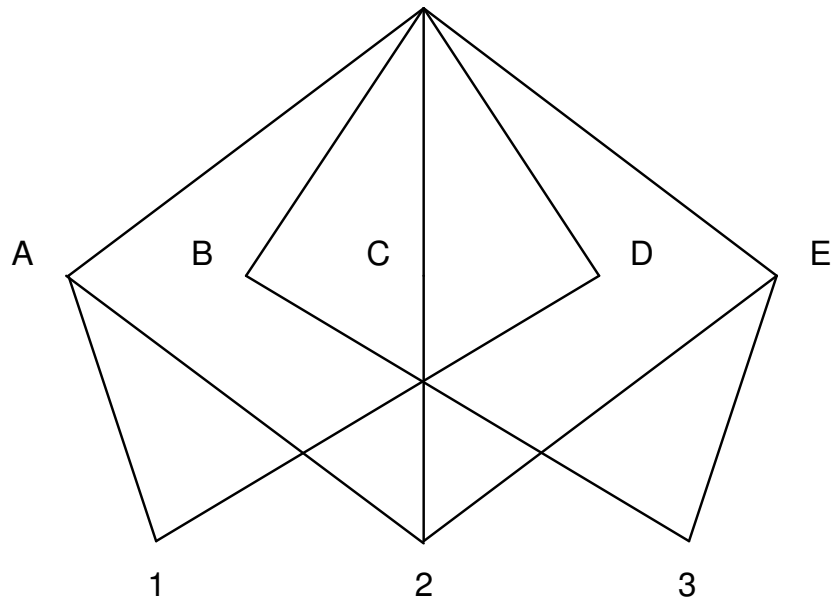


Figure 7: Cross-nested structure of the route choice example

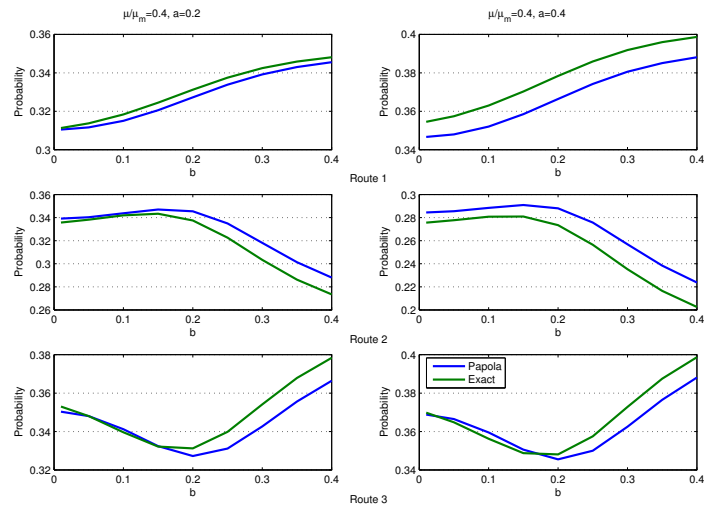


Figure 8: Comparison of choice probability between Papola's conjecture and exact formulation

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 involves 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).

However, because this procedure requires the solution of a possibly large system of non-linear equations, involving double integrals, it could involve some numerical errors in practice, and computing time may be significantly high. In such a context, Papola's conjecture can be considered as a useful practical approximation.

## 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. and Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice, in R. Hall (ed.), *Handbook of Transportation Science, Second Edition*, Kluwer, pp. 7–37.
- 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, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- Bierlaire, M. (forthcoming). A theoretical analysis of the cross-nested logit model, *Annals of operations research* . Accepted for publication.
- Bierlaire, M., Axhausen, K. and Abbay, G. (2001). Acceptance of model innovation: the case of the swissmetro, *Proceedings of the 1st Swiss Transport Research Conference*.
- Bierlaire, M., Bolduc, D. and Godbout, M.-H. (2004). An introduction to BIOGEME version 1.0. [roso.epfl.ch/biogeme](http://roso.epfl.ch/biogeme).
- Bierlaire, M. and Crittin, F. (forthcoming). Solving noisy large scale fixed point problems and systems of nonlinear equations, *Transportation Science* .
- Dennis, J. E. and Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
- Koppelman, F. and 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 A. Karlquist *et al.* (ed.), *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.
- Prashker, J. and 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. and Flannery, B. P. (2002). *Numerical Recipes in C++: The Art of Scientific Computing*, second edn, 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, *MP* **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. 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 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.