A Path-multilevel Cross Nested Logit Model for Route Choice which Allows Implicit Path Enumeration

Vittorio Marzano* Andrea Papola[†]

*Università degli Studi di Napoli "Federico II"
Via Claudio 21
80125 Naples, Italy
vmarzano@unina.it

[†] Università degli Studi di Napoli "Federico II"
Via Claudio 21
80125 Naples, Italy
papola@unina.it

1 Introduction

The simulation of route choice is characterized, in particular in an urban context, by a large amount of alternatives strongly correlated. Therefore, a route choice model should allow any covariance matrix structure to simulate the effects of path overlapping, and should be easy in its applications to large scale networks, that is it should allow implicit path enumeration and a quick calculation of choice probabilities.

The model which better answers to these properties is the Multinomial Probit model, proposed by Thurstone (1927) and particularized to route choice context by Daganzo and Sheffi (1977). Since it does not allow the calculation of choice probabilities in a closed form, some models have been proposed in the following years in order to overcome this limit, keeping Probit properties at the same time. The C-Logit model (Cascetta et al. 1996) tries to simulate the effects of correlation within a Multinomial Logit specification by introducing a route disutility factor, called commonality factor, proportional to the overlapping of that route with all the other routes in the network. Although a particular specification of the commonality factor which allows implicit path enumeration (Russo and Vitetta 2001) is available, the C-Logit model shows some problems in simulation since it does not manage to reproduce the expected choice probabilities within some test networks (Papola 2003). The Cross Nested Logit model (Small 1987, Vovsha 1997, Papola 2000, Wen and Koppelman 2001), whose application to route choice was proposed by Vovsha and Bekhor (1998) in the form of the Link Nested Logit model, is a closed-form model that seems to allow any covariance matrix. Unfortunately, a LNL implicit path enumeration algorithm can be implemented only for a particular specification of the model parameters (Prasker and Bekhor 1998), which shows some problems in simulation.

This paper proposes an original contribution to route choice simulation, which takes again and generalizes the Dial (1971) approach: the choice of a route is assumed as the result of the choices made at all the nodes of that route. In detail, we assume that in each node the user chooses between all the links exiting from that node, on the basis of their utility (typically a generalized cost) and of an inclusive utility representative of what he can find from that point of the network until the destination. Therefore, the choice probability of a path can be expressed as the product of the choice probabilities at each node of the path. It is important to underline that all other route choice models presented in the literature so far, even if quite different in their specification, calculate path probabilities under a different hypothesis: the user chooses the route alternative on the basis of its utility and of its correlation with all other routes. The proposed choice mechanism can be reproduced with the "choice graph" characteristic of a Path Multilevel Cross Nested Logit (PM-CNL) structure (Papola and Marzano 2003).

2 The model structure

The PM-CNL model is a generalization of the Cross Nested Logit model providing for a multiple nesting structure. The first Multilevel CNL model was proposed by Bierlaire (2002), who demonstrates that a GEV generating function and a corresponding random utility model can be derived correspondingly to any multilevel cross nested correlation structure (network GEV). Papola and Marzano (2003) introduce a different Multilevel CNL model, the PM-CNL model, in order to simulate the contexts of sequential choices; the model structure can be seen as a "choice graph" and the elementary alternatives are the paths of this choice graph.

The specification of a PM-CNL model requires the definition of the choice graph and of the parameters. In detail, the specification of the choice graph requires the definition of the nests and of the alternatives belonging to each nest, while the parameter specification requires the definition of the degree of membership and of the variance parameters. In order to describe the model structure, some definitions are preliminary introduced. The generic link l is determined by its starting node TL(l), named tail of the link, and its ending node HD(l), named head of the link.

Taking again Dial's definition (1971), a link l is efficient with respect to the origin o if the cost of the shortest path $Z_{o,\mathrm{TL}(l)}$ to reach its tail from origin is lower than the cost of the shortest path $Z_{o,\mathrm{HD}(l)}$ to reach its head, and it is efficient with respect to destination d if $Z_{\mathrm{HD}(l),d}<Z_{\mathrm{TL}(l),d}$. Let FSE($l$) (forward star efficient) be the set of the efficient links, both with respect to origin and destination, whose tail is the head of link l; analogously, let BSE(l) (backward star efficient) be the set of the efficient links whose head is the tail of the link l.

The root of the choice graph of the proposed model coincides with the origin o, and each efficient link l represents an intermediate node, if $\mathrm{HD}(l)\neq d$, or a final node, if $\mathrm{HD}(l)=d$, of the choice graph. The ancestors of the node representing link l are all the nodes representing the links belonging to BSE(l), and the descendants of the node representing link l are all the nodes representing the links belonging to FSE(l); in the following the generic ancestor of link l will be denoted with $a(l)$. Consequently, each efficient link of the network, except for the links entering into the destination, identifies a nest of the choice graph which all the links of its efficient forward star belong to, and which belongs to all the nests identified by the links of its efficient backward star.

At each node of the choice graph, intermediate (link l) or initial (origin o), we assume that the user chooses between all the available alternatives, formed by the links belonging respectively to FSE(l) and FSE(o) which represents themselves some intermediate or final nodes. The utility of each link l, given the ancestors $a(l)$ and $a(a(l))$, is expressed as the sum of the generalized cost c_l of the link and of an inclusive variable $Y_{l/\mathrm{a}(l)}$ representative of what can be found from the head of link l until the destination. Given this choice graph, the choice probability of the path k can be can be expressed with the PM-CNL model:

$$p[k]=\prod_{l\in L_k} p[l/a(l),a(a(l))] \tag{1}$$

where L_k is the ordered set of the links belonging to path k and the choice probability of the link l, conditioned to the choice of the two ancestors $a(l)$ and $a(a(l))$ is given by:

$$p[l/a(l),a(a(l))]=\frac{\alpha_{a(l),l}{}^{\theta_{a(l)}/\theta_{a(a(l))}}\cdot e^{-c_l/\theta_{a(l)}+\theta_l/\theta_{a(l)}\cdot Y_{l/a(l)}}}{\sum_{l'\in FSE(a(l))}\alpha_{a(l),l'}{}^{\theta_{a(l)}/\theta_{a(a(l))}}\cdot e^{-c_{l'}/\theta_{a(l)}+\theta_{l'}/\theta_{a(l)}\cdot Y_{l'/a(l)}}} \tag{2}$$

In eqn (2) $\alpha_{\mathrm{a}(l),l}$ represents the degree of nesting of link l to the group formed by the ancestor a(l) and by definition it results:

$$\sum_{a(l)\in BSE(l)}\alpha_{a(l),l}=1 \tag{3}$$

Besides, like in the Multilevel Nested Logit, the θ are variance parameters and $Y_{l/\mathrm{a}(l)}$ is the inclusive variable of the following choice levels expressed as:

$$Y_{l/a(l)} = \begin{cases} 0 & \text{if FSE}(l) = \varnothing \\ \ln \sum_{l' \in FSE(l)} \alpha_{l,l'}{}^{\theta_l / \theta_{a(l)}} e^{-c_{l'} / \theta_l + \theta_{l'} / \theta_l \cdot Y_{l'}} & \text{otherwise} \end{cases} \tag{4}$$

The equation (2) can be extended to the links $l' \in \text{FSE}(o)$ and $l'' \in \text{FSE}(l')$, which do not have two ancestors, by setting $\theta_{a(l)}=\theta_0$ for the links $l' \in \text{FSE}(o)$ and $\theta_{a(a(l))}=\theta_0$ for the links $l'' \in \text{FSE}(l')$, and considering that from the (2) it follows that $\alpha_{o,l}=1\ \forall l \in \text{FSE}(o)$. Finally, substituting equation (2) into (1) the following expression occurs:

$$p[k] = \prod_{l \in L_k} \frac{\alpha_{a(l),l}{}^{\theta_{a(l)}/\theta_{a(a(l))}} \cdot e^{-c_l/\theta_{a(l)} + \theta_l/\theta_{a(l)} \cdot Y_{l/a(l)}}}{\sum_{l' \in FSE(a(l))} \alpha_{a(l),l'}{}^{\theta_{a(l)}/\theta_{a(a(l))}} \cdot e^{-c_{l'}/\theta_{a(l)} + \theta_{l'}/\theta_{a(l)} \cdot Y_{l'/a(l)}}} \tag{5}$$

The expression of the degree of membership of link l to the group formed by the generic ancestor $a(l) \in \text{BSE}(l)$ can be expressed in the following way:

$$\alpha_{a(l),l} = \frac{\dfrac{n_{o,a(l)}}{Z_{o,TL(a(l))} + c_{a(l)}}}{\sum_{a'(l) \in BSE(l)} \dfrac{n_{o,a'(l)}}{Z_{o,TL(a'(l))} + c_{a'(l)}}} \tag{6}$$

where $Z_{o,\text{TL}(a(l))}$ is the cost of the shortest path from the origin to the tail of link $a(l)$ and $n_{o,a(l)}$ is the number of paths connecting the origin o with the tail of the link l. So, this degree of nesting increases as $n_{o,a(l)}$ increases and as $Z_{o,\text{TL}(a(l))}$ decreases, with an intuitive physical interpretation.

The variance parameters are specified following the Daganzo and Sheffy (1977) hypothesis, adapted to an homoscedastic context:

$$\text{Var}[U_k] = \xi Z_{o,d} \qquad \forall k \in od. \tag{7}$$

being $Z_{o,d}$ the shortest path between the origin o and the destination d. Therefore, from the expression of the variance for the Logit model family and from the (7) it occurs:

$$\theta_0 = \frac{\sqrt{6 \cdot \xi \cdot Z_{o,d}}}{\pi} \tag{8}$$

and for the generic link l:

$$\theta_l = \frac{\sqrt{6 \cdot \xi \cdot Z_{HD(l),d}}}{\pi} \tag{9}$$

3 The model properties: methodology and results

The numerical and operational characteristics of the proposed model have been checked through some boundary cases on toy networks, in which some expectations about route choice probabilities can be formulated by the analyst under the assumption of Papola (2003). The results show at first the PM-CNL model to be able to reproduce all the expected probability values. Furthermore, a “nesting effect” can be observed in addition to the usual “correlation effect”, as a consequence of the different hypothesis mentioned above. In other words, the model reproduces an users’ preference towards paths which allow diversions rather than paths which do not, being equal utility and correlation.

4 The implicit path enumeration algorithm

An implicit path enumeration algorithm for the stochastic network loading with the PM-CNL model is also presented. The procedure allows at first the definition of the choice graph for each origin-destination pair, through the identification of the membership degrees and of the variance parameters; then choice probabilities and link flows are determined. The structure of the algorithm is based, taking again Dial’s algorithm (1971), on a reformulation of the choice probability (2) by introducing a conditioned weight $w_{l/\mathrm{a}(l),\mathrm{a}(\mathrm{a}(l))}$ and a conditioned inclusive utility $W_{l/\mathrm{a}(l)}$.

The procedure requires, for each o-d pair, a network exploration from the destination to the origin in order to calculate the parameters and the introduced variables, and a network exploration from origin to destination for the network loading. Its computational complexity is analogous to that of the Dial’s algorithm with efficient links with respect both to the origin and destination.

5 Conclusion

This paper presents an original Path Multilevel Cross Nested Logit route choice model. The theoretical, numerical and operative aspects required for the simulation of the route choice are all satisfied; besides, it is based on a more realistic assumption with respect to the other models available in the literature. For these reasons, the authors propose such model as a benchmark for route choice simulation.

References

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

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

E. Cascetta (2001). *Transportation system engineering: theory and methods*. Kluwer ed.

E. Cascetta, A. Nuzzolo, F. Russo, A. Vitetta (1996). "A modified Logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks". Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Jean-Baptiste Lesort ed., Pergamon.

C.F. Daganzo, Y. Sheffi (1977). "On stochastic models of traffic assignment". *Transportation Science*, Vol. 11, n° 3.

A. Daly, S. Zachary (1976). *Improved Multiple Choice Models*. Mimeo

R.B. Dial (1971). "A probabilistic multipath traffic assignment model which obviates path enumeration". *Transportation Research*, No. 5, pp. 83-111.

A. Papola (2000). "Some developments of the cross-nested logit model". Proceedings of the 9th IATBR Conference.

A. Papola (2003). "Applicabilità del modello Link nested logit alla simulazione della scelta del percorso". Proceedings of "Metodi e Tecnologie dell'Ingegneria dei Trasporti", Reggio Calabria 2001, Franco Angeli ed.

A. Papola, V. Marzano (2003). "Il modello Cross Nested Logit multi livello". *Internal report*
J.N. Prashker, S. Bekhor (1998). "Investigation of stochastic network loading procedures". Proceedings of the 77th Annual Meeting of the Transportation Research Board.

F. Russo, A. Vitetta (2001). "Un modello di scelta del percorso che affronta i problemi di enumerazione esplicita e sovrapposizione dei percorsi". Proceedings of "Metodi e Tecnologie dell'Ingegneria dei Trasporti", Reggio Calabria 1999, Franco Angeli ed.

K. Small (1987). "A discrete choice model for ordered alternatives". *Econometrica* n. 55, pp. 409-424

L. Thurstone (1927). "A law of comparative judgement". *Psycological Review* n. 34, pp. 273-286

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

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

C. H. Wen, F. S. Koppelman (2001). "The generalized nested logit model". *Transportation Research* B n. 35(8) pp. 627-641.

H.C.L. Williams (1977). "On the formation of travel demand models and economic evaluation measures of user benefit". *Environment and Planning* A, pp. 285-344.