11-2016

A method of integrating correlation structures for a generalized recursive route choice model

Tien MAI

Singapore Management University, atmai@smu.edu.sg

Transportation Research Part B 93 (2016) 146–161

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A method of integrating correlation structures for a generalized recursive route choice model

Tien Mai*

Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal and CIRRELT, Canada

ARTICLE INFO

Article history:
Available online 30 July 2016

Keywords:
Recursive network MEV
Contraction mapping
Integrated network
Recursive cross-nested
Value iteration
Maximum likelihood estimation
Cross-validation

ABSTRACT

We propose a way to estimate a generalized recursive route choice model. The model generalizes other existing recursive models in the literature, i.e., (Fosgerau et al., 2013b; Mai et al., 2015c), while being more flexible since it allows the choice at each stage to be any member of the network multivariate extreme value (network MEV) model (Daly and Bierlaire, 2006). The estimation of the generalized model requires defining a contraction mapping and performing contraction iterations to solve the Bellman's equation. Given the fact that the contraction mapping is defined based on the choice probability generating functions (CPGF) (Fosgerau et al., 2013b) generated by the network MEV models, and these CPGFs are complicated, the generalized model becomes difficult to estimate. We deal with this challenge by proposing a novel method where the network of correlation structures and the structure parameters given by the network MEV models are integrated into the transport network. The approach allows to simplify the contraction mapping and to make the estimation practical on real data.

We apply the new method on real data by proposing a recursive cross-nested logit (RCNL) model, a member of the generalized model, where the choice model at each stage is a cross-nested logit. We report estimation results and a prediction study based on a real network. The results show that the RCNL model performs significantly better than the other recursive models in fit and prediction.

1. Introduction

Given a transport network with links and nodes, and given a origin-destination pair, the route choice problem deals with identifying which route that a traveler would take. Discrete choice models are generally used in the context. There are two main issues associated with the use of discrete choice models, namely, choice sets of paths are unknown to the analyst, and path utilities may be correlated. We consider a generalized recursive route choice model where the choice at each stage is a family of multivariate extreme value (MEV) models. The model has the advantages from the existing recursive models (Fosgerau et al., 2013a; Mai et al., 2015c), as it can be consistently estimated and is easy for prediction without sampling of choice sets. It is, moreover, flexible in the sense that it allows to capture the correlation at each choice stage by many convenient static discrete choice models, e.g., multinomial, nested logit (Ben-Akiva, 1973), cross-nested logit (Vovsha and Bekhor, 1998) or multi-level cross-nested models (Daly and Bierlaire, 2006). However, the generalized model is impractical to estimate if using the estimation methods proposed for the other recursive models. We therefore propose an innovative

* Corresponding author.
E-mail address: anh-tien.mai@polymtl.ca

http://dx.doi.org/10.1016/j.trb.2016.07.016
0191-2615/

method that allows to integrate the correlation structure given by the model at each stage into the transport network and simplify the estimation.

The recursive logit (RL) model proposed by Fosgerau et al. (2013a) can be consistently estimated on real data without sampling any choice sets of paths. It is assumed that travelers choose states (nodes or links) in a sequential manner. At each state they maximize the sum of the random utility associated with a successor state (instantaneous utility) and the expected maximum utility from the state to the destination (also known as the value functions). The random terms of the instantaneous utilities are assumed to be independently and identically distributed (i.i.d.) extreme value type I and the RL model is equivalent to a multinomial logit (MNL) model over choice sets of all feasible path alternatives. The RL model hence inherits the independence of irrelevant alternatives (IIA) property which is undesirable in a route choice setting (for instance Mai et al., 2015d). Recently, Mai et al. (2015c) proposed the nested RL (NRL) model that relaxes the IIA property of the RL model by assuming that scale parameters are link specific. Both the RL and NRL models, however, assume that the choice at each stage are MNL, so the correlation between the utilities of successor states cannot be captured.

In this paper, we consider a generalization of these models where the choice model at each stage can be any member of the network MEV model (Daly and Bierlaire, 2006), e.g., the MNL, the nested logit (Ben-Akiva, 1973) or cross-nested logit (Vovsha and Bekhor, 1998). This model, called the recursive network MEV (RNMEV), is fully flexible in the sense that the cross-nested logit can approximate any additive random utility model (for instance Fosgerau et al., 2013b). However, the estimation of the generalized model leads to a complicated dynamic programming (DP) problem, which is cumbersome to solve. We explain this challenge in more detail in the following.

All the existing recursive route choice models are based on the dynamic discrete choice framework proposed by Rust (1987), where the estimation requires solving a Bellman's equation to obtain the value functions. Rust (1987) show that the value functions can be computed by defining a contraction mapping $V = \mathcal{T}_\beta(V)$, where $\mathcal{T}_\beta$ is a contraction mapping associated with parameters β, and V is the vector of the value functions. The fixed point solution, or the value functions, then can be obtained by using the method of successive approximations (or value iteration) based on $\mathcal{T}_\beta$. In practice, this method needs to be formulated as matrix operations in order to deal with problems with large number of states. In the RL and NRL models, due to the simple choice probability generating functions (CPGF) (Fosgerau et al., 2013b) given by the MNL model at the choice stages, the corresponding contraction mappings have simple forms (even linear form in the RL model). Consequently, the corresponding contraction iterations can be formulated easily as matrix operations, and the RL and NRL models can be estimated quickly using real networks. On the contrary, the CPGFs given by the network MEV model are complicated. They have closed forms but need to be computed recursively based on the network of correlation structure (see for instance Daly and Bierlaire, 2006). Hence, the contraction mapping given by the RNMEV model is complicated, and it is not straightforward to formulate the contraction iterations as matrix operations. This explains why the estimation of the RNMEV model is extremely expensive or even impractical if using directly the standard approaches proposed in Fosgerau et al. (2013a); Mai et al. (2015c) and Rust (1987). In this context, we note that the generalized model also relaxes the well-known CLOGIT assumption from the Rust's model (Rust, 1987), i.e., the assumption that the choice at each stage is MNL.

We propose an innovative method that allows us to simplify the contraction mapping given by the RNMEV and to quickly estimate the generalized model on real data. More precisely, we consider the networks of correlation structures given by the network MEV models (also known as MEV-networks) at the choice stages. We create a new artificial network by integrating the MEV-networks into the transport network and associate the states and arcs of the new network with new parameters and utilities based on those from the original network and the structure parameters of the network MEV models. We show similarities between the value functions given by the RNMEV model and those given by the NRL model on the integrated network. This method allows us to avoid recursively computing the CPGFs given by the network MEV models, and to use the contraction mapping from the NRL model to obtain the value functions of the RNMEV model. This approach, therefore, greatly simplifies the estimation of the RNMEV model.

We apply the new method on real data by proposing a recursive cross-nested logit (RCNL) model, a member of the RNMEV model, where the choice model at each stage is a cross-nested logit model (Vovsha and Bekhor, 1998). In the RCNL model, the variance-covariance matrix at each choice stage is no longer diagonal as in the RL and NRL models. This model therefore exhibits a more general correlation structure, compared to the other recursive models. We provide estimation and prediction results for a real network, and show that the RCNL model can be estimated in reasonable time while performing significantly better than the RL and NRL models in fit and prediction.

This paper makes two main contributions. First, we propose the method of integrating correlation structures to simplify the estimation of the RNMEV. Second, we propose the RCNL model that can flexibly capture the correlation structure at each choice stage, and we provide estimation and cross-validation results of the RCNL for a real network using real observations. Moreover, the estimation code for estimating the RCNL is available in an open source project and we share it freely upon request.

The paper is structured as follows. Section 2 introduces the RNMEV model and Section 3 presents the method of integrating correlation structure. Section 4 discusses in detail the maximum likelihood estimation. Section 5 presents the RCNL model. We provide the estimation results and a cross-validation study in Section 6, and finally, Section 8 concludes. The proofs of the theorems presented in this paper are provided in Appendixes A and B.

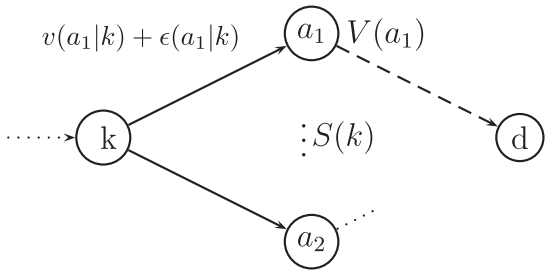

Fig. 1. Illustration of notation.

2. Generalized recursive route choice model

In the recursive route choice models proposed in Fosgerau et al. (2013a) and Mai et al. (2015c), the path choice problem is formulated as a sequence of link choices and modeled in a dynamic discrete choice framework. At each sink node of a link the decision maker chooses the utility-maximizing outgoing link with link utilities given by the instantaneous utility and the expected maximum utility to the destination. The random terms at each state are i.i.d. extreme value type I, so the choice model at each stage is MNL. In this section, we generalize the RL and NRL models by assuming the choice at each stage to be a network MEV model (Daly and Bierlaire, 2006). In order to better describe the generalized model, we consider the road network as a set of states and arcs connecting states. The states can be nodes in the network, or links as in Mai et al. (2015c) and Fosgerau et al. (2013a). The instantaneous utilities are defined for states conditional on other states and the path choice problem is formulated as a sequence of state choices, and there are states representing destinations in the road network.

A directed connected graph (not assumed acyclic) $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ is considered, where $\mathcal{S}$ and $\mathcal{A}$ are the sets of states and arcs, respectively. For each state $k \in \mathcal{A}$, we denote the set of successor states of k by $S(k)$ (if states are links, $S(k)$ is the set of the outgoing links from the sink node of k). Moreover, we associate an absorbing state with the destination of a given individual by extending the network with a dummy state d that has no successor (see Fig. 1). The set of all states is therefore $\tilde{\mathcal{S}} = \mathcal{S} \cup \{d\}$ and the corresponding deterministic utility is $v(d|k) = 0$ for all k that connects to d.

Given two states $a, k \in \tilde{\mathcal{S}}$ and individual n, the following instantaneous utility is associated with state a conditional on k

$$u^n(a|k;\beta) = v^n(a|k;\beta) + \epsilon(a|k;\beta) - \frac{\gamma}{\mu_k(\beta)},\ \forall k \in \mathcal{S}, a \in S(k),$$

where β is a vector of parameters to be estimated, γ is the Euler's constant and random terms $(\epsilon(a|k),\ a \in S(k))$ follow an MEV distribution, with the CPGF G_k of homogeneous degree $\mu_k > 0$ generated by the network MEV model. We note that the term $\frac{\gamma}{\mu_k}$ is used in order to ensure that the random term has zero mean. The deterministic term $v^d(a|k),\ a \in S(k)$, is assumed negative for all states except for the dummy d that equals 0, i.e., $v^d(d|k) = 0,\ \forall k \in \mathcal{S}$. For notational simplicity, we omit from now on an index for individual n but note that the utilities can be individual specific. Given a state $k \in \mathcal{S}$, the next state is chosen by taking the maximum utility as

$$\text{argmax}_{a \in S(k)} \left\{ v(a|k;\beta) + V^d(a;\beta) + \epsilon(a|k;\beta) - \frac{\gamma}{\mu_k(\beta)} \right\},\ \forall k \in \mathcal{S},$$

where $V^d(a;\beta),\ \forall a \in \tilde{\mathcal{S}}$, is the expected maximum utility (or value function) from the state a to the destination, which is recursively defined by Bellman's equation

$$V^d(k;\beta) = \mathbb{E}\left[\max_{a \in S(k)} \left(v(a|k;\beta) + V^d(a;\beta) + \epsilon(a|k;\beta) - \frac{\gamma}{\mu_k(\beta)} \right) \right],\ \forall k \in \mathcal{S}, \tag{1}$$

and $V^d(d) = 0$. The superscript d indicates that the value functions are destination specific and they also depend on parameters β. However, for notational simplicity we omit from now on β and superscript d from the value functions $V()$ and the utilities $v()$. According to McFadden et al. (1978), the Bellman's equation can be written as

$$V(k) = \frac{\ln G_k\left(e^{v(a|k)+V(a)}, a \in S(k) \right)}{\mu_k},\ \forall k \in \mathcal{S}, \tag{2}$$

and note that $V(d) = 0$. If we define a vector of size $|\tilde{\mathcal{S}}|$ ($|\cdot|$ is the cardinality operator) with entries

$$Y_k = e^{\mu_k V(k)},\ \forall k \in \tilde{\mathcal{S}},$$

then the system in (1) can be written as

$$Y_k = \begin{cases} G_k(e^{v(a|k)} Y_a^{1/\mu_a}, a \in S(k)),\ \forall k \in \mathcal{S} & \text{if } k \in \mathcal{S} \\ 1 & \text{if } k = d \end{cases}. \tag{3}$$

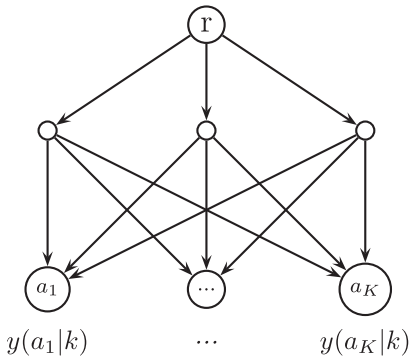

Fig. 2. A network of correlation structure at state k.

Moreover, the probability of choosing state a given state k is given by the MEV model

$$P(a|k) = \delta(a|k)\frac{y(a|k)\frac{\partial G_k}{\partial y(a|k)}(y(a'|k), a' \in S(k))}{\mu_k G(y(a'|k), a' \in S(k))}, \ \forall k, a \in \tilde{\mathcal{S}}, \tag{4}$$

where $y(a|k) = e^{v(a|k)}Y_a^{1/\mu_a}$, $\forall a \in S(k)$. Note that we include $\delta(a|k)$ that equals one if $a \in S(k)$ and zero otherwise so that the probability is defined for all states $a, k \in \tilde{\mathcal{S}}$.

The probability of a path σ defined by a sequence of states $\sigma = [k_0, k_1, \ldots, k_I]$ has a more complicated form than the ones given by the RL and NRL models. In general, it can be expressed as

$$P(\sigma) = \prod_{i=0}^{I-1} P(k_{i+1}|k_i), \tag{5}$$

in which $P(k_{i+1}|k_i)$ can be computed using (4).

Now we turn our attention to the network MEV model at each stage. We assume that for each state $k \in \mathcal{S}$, the respective CPGF $G_k(y)$ is generated by a network MEV model based on a cycle-free network of correlation structure $\mathcal{G}_k = (\mathcal{S}_k, \mathcal{A}_k, C_k)$, where $\mathcal{S}_k$ is the set of states, $\mathcal{A}_k$ is the set of arcs and C_k is the set of states that represent alternatives. Note that the set of states representing alternatives in this network MEV model is also the set of next states from k that is $S(k)$. So each state $i \in C_k$ corresponds to only one state $a \in S(k)$ and vice versa. Fig. 2 illustrates a network of correlation structure at state k. Each arc $(i, j) \in \mathcal{A}_k$ is associated with a positive parameter α_{ij}^k and each state $i \in \mathcal{S}_k$ is associated with a positive scales ξ_i^k. The CPGF (with respect to a vector of parameters y) associated with each state in $\mathcal{S}_k$ are defined as

$$G_k^i(y) = y_i^{\xi_i^k}, \quad i \in C_k, \tag{6}$$

and

$$G_k^i(y) = \sum_{j \in S_k(i)} \alpha_{ij}^k (G_k^j(y))^{\xi_i^k/\xi_j^k}, \forall i \in \mathcal{S}_k \backslash C_k, \tag{7}$$

where $S_k(i)$ is the set of the successors of state i in network $\mathcal{G}_k$.

We obtain the CPGF $G_k(y)$ as $G_k(y) = G_k^r(y)$, where r is the root of network $\mathcal{G}_k$. Daly and Bierlaire (2006) show that $G_k(y) = G_k^r(y)$ is a ξ_r^k-MEV CPGF and μ_k is the homogeneous degree of $G_k(y)$, so $\xi_r^k = \mu_k$. Moreover, Daly and Bierlaire (2006) show that the probability $P_k(i|C_k; y)$ of choosing alternative $i \in C_k$ can be expressed based on the CPGFs defined in (6) and (7) as

$$P_k(i|C_k; y) = \sum_{[j_0, \ldots, j_I] \in \Omega^k(i)} \prod_{t=0}^{I-1} \frac{\alpha_{j_t j_{t+1}}^k (G_k^{j_{t+1}}(y))^{\xi_{j_t}^k/\xi_{j_{t+1}}^k}}{G_k^{j_t}(y)}, \tag{8}$$

where $\Omega^k(i)$ is the set of all paths connecting the root r and i. A path is defined by a sequence of states $[j_0, \ldots, j_I]$ such that $j_{t+1} \in S_k(j_t)$, $\forall t = 0, \ldots I-1$,where j_0 is the root r and j_I represents alternative i. If we denote y_k a vector of size $|S(k)|$ with entries $(y_k)_a = y(a|k) = e^{v(a|k)}Y_a^{1/\mu_a}$, $\forall a \in S(k)$, then according to (3) we have $Y_k = G_k(y_k)$ and the probability $P(a|k)$ for a state $a \in S(k)$ can be computed by using (8). In other words, $P(a|k) = P_k(i_a|C_k; y_k)$, where i_a is a state in C_k corresponding to state $a \in S(k)$.

The network MEV model generalizes many MEV models in the literature and examples are the MNL, the nested logit (Ben-Akiva, 1973), the paired combinatorial logit (Koppelman and Wen, 2000), the generalized nested logit (Wen and Koppelman, 2001), the ordered MEV model (Small, 1987), the cross-nested logit model (Vovsha and Bekhor, 1998) and the GenL

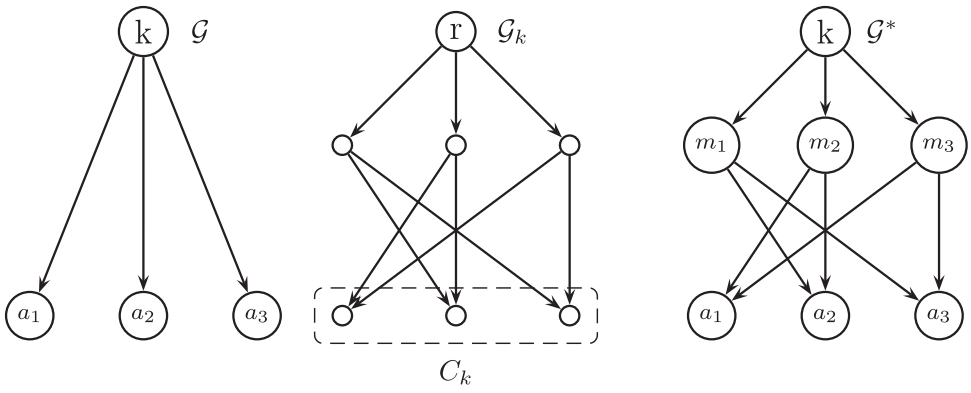

Fig. 3. Original and integrated network.

model (Swait, 2001) models. So the RNMEV model allows to capture the correlation at each choice stage by many convenient discrete choice models. Indeed, if all the G_k functions, $\forall k \in \mathcal{S}$, refer to the MNL model, then the RNMEV becomes the NRL model.

The estimation of the RNMEV model requires solving the Bellman's equation in (2). This can be done by defining a contraction mapping $V = \mathcal{T}_\beta(V)$ (Rust, 1987), and applying the method of successive approximations (value iteration), i.e., performing $V^{(t+1)} = \mathcal{T}_\beta(V^{(t)})$ iteratively until the fixed point solution is found. The contraction mapping $\mathcal{T}_\beta$ is defined based on CPGFs $G_k(\cdot)$, $\forall k \in \mathcal{S}$, where the computation of CPGF $G_k(\cdot)$ requires performing the recursive equations in (7) and (6), leading to the fact that $\mathcal{T}_\beta$ is complicated. Consequently, it is not straightforward to formulate the method of successive approximations as matrix operations, and the computation of the value functions becomes impractical if using standard approaches. The gradients of the value functions are cumbersome to evaluate as well. In the next section, we show how to simplify the computation of the value functions as well as the choice probabilities by integrating the networks of correlation structures $\mathcal{G}_k$, $\forall k \in \mathcal{S}$, into the road network $\mathcal{G}$.

3. Method of integrating correlation structures

In this section, we show how to integrate the MEV-networks at choice stages into the transport network in order to simplify the contraction mapping given by the RNMEV model. More precisely, we first introduce a method to integrate the networks of correlation structures $\mathcal{G}_k$, $\forall k \in \mathcal{S}$, into the road network $\mathcal{G} = (\mathcal{S}, \mathcal{A})$, second we show similarities between the RNMEV and the NRL model on the integrated network. These similarities allow us to use the contraction mapping given by the NRL to quickly compute the value functions of the RNMEV model.

3.1. Integrated network

Given a state $k \in \mathcal{S}$, the choice at k is a network MEV model based on a network or correlation structure $\mathcal{G}_k = (\mathcal{S}_k, \mathcal{A}_k, C_k)$. As mentioned earlier, the set of states C_k (representing alternatives) is also the set of next states from k, i.e., $S(k)$. So in order to integrate $\mathcal{G}_k$ to the road network we assume that $C_k \equiv S(k)$ and $k \equiv r$. Hence, the integrated network $\mathcal{G}^* = (\mathcal{S}^*, \mathcal{A}^*)$ can be created by adding all sets $\mathcal{S}_k$ and $\mathcal{A}_k$, $\forall k \in \mathcal{S}$, to the set of states $\mathcal{S}$ and set of arcs $\mathcal{A}$. In other words

$$\mathcal{S}^* = \bigcup_{k \in \mathcal{S}} \mathcal{S}_k \text{ and } \mathcal{A}^* = \bigcup_{k \in \mathcal{S}} \mathcal{A}_k. \tag{9}$$

Basically, the new network $\mathcal{G}^*$ is created by adding new states to $\mathcal{G}$. For each $k \in \mathcal{S}$, we add new states such that the subnetwork between k and all states $a \in S(k)$ is similar to the network of correlation structure $\mathcal{G}_k$. We also denote $S^*(k)$ as the set of successor states of state k in network $\mathcal{G}^*$.

We note that, due to the properties of the MEV-networks (for instance Daly and Bierlaire, 2006), $\mathcal{G}^*$ remains connected and there are paths connecting between any two states $k, a \in \tilde{\mathcal{S}}$, $a \in S(k)$. For the sake of illustration, we show in Fig. 3 a small example where state k has three successors a_1, a_2 and a_3 as illustrated in the left part of Fig. 3. The network of correlation structure $\mathcal{G}_k$ is given in the middle of the figure and in the right we show the integrated network $\mathcal{G}^*$ at state k.

We introduce the following proposition related to the properties of the network $\mathcal{G}^*$, which are easy to verify.

Proposition 1. *Network $\mathcal{G}^*$ has the following properties*

(i) *Given a state $i \in \mathcal{S}^*$, there is a state $k \in \mathcal{S}$ such that $i \in \mathcal{S}_k$.*
(ii) *Given a state $i \in \mathcal{S}^*$, if $i \notin \tilde{\mathcal{S}}$ then there exists only one state $k \in \mathcal{S}$ such that $i \in \mathcal{S}_k$.*
(iii) *Given a state $i \in \mathcal{S}_k$, if $i \in \tilde{\mathcal{S}}$ then $i = k$ or $i \in S(k)$.*

(iv) *Given a state* $i \in \mathcal{S}_k$, *if* $i \notin C_k$ *then* $S_k(i) = S^*(i)$.
(v) $\tilde{\mathcal{S}} \cap \mathcal{S}_k = \{k\} \cup S(k)$ *and* $|\mathcal{S}^*| = |\tilde{\mathcal{S}}| + \sum_{k\in\mathcal{S}} (|\mathcal{S}_k| - |S(k)| - 1)$.
(vi) $\mathcal{A}_k \cap \mathcal{A}_h = \emptyset$ $\forall k, h \in \mathcal{S}, k \neq h$ *and* $|\mathcal{A}^*| = \sum_{k\in\mathcal{S}} |\mathcal{A}_k|$.

Proof. (i), (iii), (v) and (vi) are obviously verified by the definition in (9). We have the fact that given a state i, if $i \in \mathcal{S}_k \cap \mathcal{S}_h$ (with $k, h \in \mathcal{S}, k \neq h$) then $i \in \tilde{\mathcal{S}}$. So if state $i \notin \tilde{\mathcal{S}}$ then there is only one state $k \in \mathcal{S}$ such that $i \in \mathcal{S}_k$. This proves (ii).

For proving (iv) we note that given $i \in \mathcal{S}_k \backslash C_k$ and if $j \in S_k(i)$ for a given $j \in \mathcal{S}^*$ then $(i, j) \in \mathcal{A}_k$, meaning that $(i, j) \in \mathcal{A}^*$ by (9), or equivalently $j \in S^*(i)$. Moreover, if $j \in S^*(i)$ then $(i, j) \in \mathcal{A}^*$. From (v) there is only a state $k' \in \mathcal{S}$ such that $(i, j) \in \mathcal{A}_{k'}$. Moreover, (ii) leads to the fact that $k = k'$, so $(i, j) \in \mathcal{A}_k$ or $j \in S_k(i)$. Finally, we obtain $S_k(i) \subset S^*(i)$ and $S^*(i) \subset S_k(i)$. Hence, $S_k(i) = S^*(i)$ and (iv) is proved. □

Now we integrate the structure parameters given by the network MEV models into the integrated network. We consider network $\mathcal{G}^* = (\mathcal{S}^*, \mathcal{A}^*)$. We note that $d \in \mathcal{S}^*$ and $S^*(d) = \emptyset$. We associate each state $i \in \mathcal{S}^*$ a positive parameter μ_i^* as

$$\mu_i^* = \begin{cases} \mu_i & \text{if } i \in \tilde{\mathcal{S}} \\ \xi_i^k & \text{if } i \notin \tilde{\mathcal{S}}, i \in \mathcal{S}_k, k \in \mathcal{S} \end{cases}. \tag{10}$$

Recall $\tilde{\mathcal{S}} = \mathcal{S} \cup \{d\}$. Note that due to Proposition 1(ii), for $i \notin \tilde{\mathcal{S}}$, there is only one set $\mathcal{S}_k$ such that $i \in \mathcal{S}_k$, so there is only one value ξ_i^k such that $i \in \mathcal{S}_k$, $k \in \mathcal{S}$.

For each arc $(i, j) \in \mathcal{A}_k, k \in \mathcal{S}$, the following deterministic utility is associated with state j conditional on i

$$v^*(j|i) = \begin{cases} \frac{\ln \alpha_{ij}^k}{\mu_i^*} & \text{if } j \notin S(k) \\ \frac{\ln \alpha_{ij}^k}{\mu_i^*} + v(j|k) & \text{if } j \in S(k) \end{cases}, \tag{11}$$

here we recall that $v(j|k)$, $k \in \mathcal{S}, j \in S(k)$, is a deterministic utility associated with state j conditional on k and α_{ij}^k are positive parameters of the network MEV model at state k.

3.2. NRL model on the integrated network

Now we apply the NRL model (Mai et al., 2015c) to network $\mathcal{G}^*$. Given two states $k, a \in \mathcal{S}^*$, $a \in S^*(k)$, the following instantaneous utility associated with state a given k

$$u^*(a|k) = v^*(a|k) + \frac{\epsilon(a) - \gamma}{\mu_k^*}, \tag{12}$$

where $\epsilon(a)$ are i.i.d extreme value type I and $v*(a|k)$, μ_k^* are defined in (11) and (10). We notice that, in order to be consistent with the RNMEV model, the scales of the random terms in the NRL model are $\frac{1}{\mu_k^*}$ instead of μ_k^* in Mai et al. (2015c).

The expected maximum utility from the sink node of k, $k \in \mathcal{S}^*$, to the destination is the value function $V^*(k)$ that is recursively defined by the Bellman's equation

$$V^*(k) = \mathbb{E}\left[\max_{a\in S^*(k)} \left\{v^*(a|k) + V^*(a) + \frac{\epsilon(a) - \gamma}{\mu_k^*}\right\}\right], \tag{13}$$

or equivalently (by the logsum)

$$\mu_k^* V^*(k) = \ln\left(\sum_{a\in S^*(k)} e^{\mu_k^*(v^*(a|k)+V^*(a))}\right), \forall k \in \mathcal{S}^* \backslash \{d\}, \tag{14}$$

and note that $V^*(d) = 0$. If we define a vector Y^* of size $|\mathcal{S}^*|$ with entries $Y_k^* = e^{\mu_k^* V^*(k)}$, $\forall k \in \mathcal{S}^*$, then the Bellman's equation becomes

$$Y_k^* = \begin{cases} \sum_{a\in S^*(k)} e^{\mu_k^* v^*(a|k)} (Y_a^*)^{\mu_k^*/\mu_a^*} & \text{if } k \neq d \\ 1 & \text{if } k = d \end{cases}. \tag{15}$$

Moreover, the probability of choosing state a given k is given by the MNL as

$$P^*(a|k) = \delta^*(a|k) \frac{e^{\mu_k^* v^*(a|k)} (Y_a^*)^{\mu_k^*/\mu_a^*}}{Y_k^*}, \forall k, a \in \mathcal{S}^*, \tag{16}$$

where $\delta^*(a|k)$ equals one if $a \in S^*(k)$ and zero otherwise so that the probability is defined for all $k, a \in \mathcal{S}^*$. We note that the system in (15) is non-linear but can be solved quickly for a large network using the approach proposed in Mai et al. (2015c), namely a value iteration method with dynamic accuracy.

3.3. Similarities between the NRL on network $\mathcal{G}^*$ and RNMEV on network $\mathcal{G}$

This section presents similarities between the value functions and choice probabilities given by the NRL on network $\mathcal{G}^*$ and those given by the RNMEV on network $\mathcal{G}$. We first introduce a theorem related to the value functions.

Theorem 1. *If vector Y^* is a solution to the non-linear system (15) then*

$$Y_k^* = G_k\left(e^{v(a|k)}(Y_a^*)^{1/\mu_a}, a \in S(k)\right), \ \forall k \in \mathcal{S}. \tag{17}$$

In other words, $Y_k = Y_k^*$, $\forall k \in \tilde{\mathcal{S}}$, is a solution to the Bellman's equation of the RNMEV given by (3).

The next theorem shows that the state choice probabilities under the RNMEV model can be expressed via the probabilities given by the NRL model on network $\mathcal{G}^*$.

Theorem 2. *If vector Y^* is a solution to the non-linear system (15), and if $Y_k = Y_k^*$, $\forall k \in \tilde{\mathcal{S}}$, then*

$$P(a|k) = \sum_{[a_0,\dots,a_I]\in\Omega^k(a)} \prod_{t=0}^{I-1} P^*(a_{t+1}|a_t), \quad \forall k \in \mathcal{S}, a \in S(k). \tag{18}$$

We recall that $\Omega^k(a)$ is the set of sequences of states connecting k and a: $[k = a_0, \dots, a_I = a]$ such that $a_t \in S_k(a_{t-1})$, $\forall t = 1, \dots, I$, and $a_{t+1} \in S_k(a_t)$, $\forall t = 0, \dots I-1$.

We provide the proofs of the two theorems in Appendixes A and B. These theorems indicate that the value functions and choice probabilities in the RNMEV model can be computed by using the respective values from the NRL model applying on network $\mathcal{G}^*$. This means that the methods proposed in Mai et al. (2015c) can be used to estimate the RNMEV model. In the next section we discuss in detail the maximum likelihood estimation.

4. Maximum likelihood estimation

The nested fixed point algorithm (Rust, 1987) is popularly used to estimate a dynamic discrete choice model. This algorithm combines an outer iterative non-linear optimization algorithm for searching over the parameter space with an inner algorithm for solving the Bellman's equation to obtain the value functions. We have shown that the computation of the value functions and choice probabilities given by the RNMEV model can be simplified by applying the NRL model to the integrated network. The NRL model can be estimated efficiently using the methods in Mai et al. (2015c), namely, estimating the value functions with dynamic accuracy, and computing the gradients of the log-likelihood function by solving systems of linear equations. In the following we briefly describe the computations of these values in the NRL model based on the integrated network.

4.1. Computation of the value functions

The main challenge associated with the NRL model is to solve the large-scale system of non-linear to obtain the value functions. Similarity to Mai et al. (2015c), we define a matrix $M^*(|\mathcal{S}^*| \times |\mathcal{S}^*|)$ with entries

$$M_{ka}^* = \delta^*(a|k)e^{\mu_k^* v^*(a|k)}, \quad \forall k, a \in \mathcal{S}^*, \tag{19}$$

and a matrix X of size $|\mathcal{S}^*| \times |\mathcal{S}^*|$ with entries

$$X(Y^*)_{ka} = (Y_a^*)^{\mu_k^*/\mu_a^*}, \quad \forall k, a \in \mathcal{S}^* \tag{20}$$

and b is a vector of size $|\mathcal{S}^*|$ with zero values for all states except for the dummy state d with a value of 1. The Bellman's equation in (15) can be written in a matrix form as

$$Y^* = [M^* \circ X(Y^*)]e + b, \tag{21}$$

where e is a vector of size $(|\mathcal{S}^*|)$ with value one for all states, and $\circ$ is the element-by-element product. A value iteration method can be used to solve this system, i.e., we start with a initial vector $(Y^*)^0$ and then for each iteration i we compute a new vector $(Y^*)^{i+1} \leftarrow [M^* \circ X((Y^*)^i)]e + b$, and iterate until a fixed point solution is found using $||(Y^*)^{i+1} - (Y^*)^i|| \le \gamma$, for a given threshold $\gamma > 0$ as stopping criteria. Mai et al. (2015c) show that the value iteration can be improved by using dynamic accuracy. The choice of initial vector is also important for the rate of convergence. Mai et al. (2015c) use the solution to the system of linear equations corresponding to the RL model (Fosgerau et al., 2013a) which is fast to compute. This choice can be improved by taking into account the solution of the previous iteration of the outer optimization algorithm (the algorithm for searching over the parameters space) and using a switching approach to select the best initial vector. More precisely, at iteration $t-1$ of the outer algorithm, we suppose that the fixed point solution is $\overline{(Y^*)}^{t-1}$. At the next iteration t, we suppose $(Y^*)^0$ is the fixed point solution of the RL model. For solving the value functions, the initial vector for the inner algorithm can be chosen by considering

$$err = ||(Y^*)^0 - [M^* \circ X((Y^*)^0)]e + b|| - ||\overline{(Y^*)}^{t-1} - [M^* \circ X(\overline{(Y^*)}^{t-1})]e + b||.$$

If $err < 0$ than vector $(Y^*)^0$ is chosen, otherwise we select $\overline{(Y^*)}^{t-1}$. This switching approach allows to select the better initial vector (closer to the fixed point solution).

4.2. Estimation

Now we derive the log-likelihood (LL) function of the RNMEV model. The LL function defined over the set of path observations $n = 1, \ldots, N$ is

$$LL(\beta) = \sum_{n=1}^{N} \ln P(\sigma_n, \beta) = \sum_{n=1}^{N} \sum_{t=0}^{I_n} \ln P(k_{t+1}^n | k_t^n). \tag{22}$$

We note that the path observations are defined based on states in network $\mathcal{G}$. Each probability $P(k_{t+1}^n | k_t^n)$, $k_{t+1}^n, k_t^n \in \mathcal{S}$, can be computed using the results of Theorems 1 and 2.

For the maximum likelihood estimation, the network $\mathcal{G}_k$, $\forall k \in \mathcal{S}$, generates MEV models with many parameters. That is ξ_i^k and α_{ij}^k, $\forall i, j \in \mathcal{S}_k, j \in S_k(i)$. Daly and Bierlaire (2006) show that the constraints $\xi_i^k \leq \xi_j^k$, $\forall i, j \in \mathcal{S}_k, j \in S_k(i)$, need to be satisfied in order to ensure that the choice at state k is consistent with McFadden's MEV theory. Based on the definition in (10) the constraints can be written as

$$\mu_i^* \leq \mu_j^*, \forall i, j \in \mathcal{S}^*, j \in S^*(i) \backslash \tilde{S}.$$

Moreover, as suggested by Daly and Bierlaire (2006), a normalization for parameters α_{ij}^k for the network MEV model at state k would require as

$$\sum_{\substack{i \in \mathcal{S}^* \\ s.t. j \in S^*(i)}} (\alpha_{ij}^k)^{\xi_i^k / \xi_j^k} = 1, \forall j \in \mathcal{S}_k, k \in \mathcal{S}.$$

This normalization, however, remains to be analyzed further.

Efficient nonlinear techniques for the problem require analytical derivatives of the LL function. The gradients of the LL function are complicated, but can be easily derived based on (22), (2) and (16). They require the gradients of Y_k^*, $\forall k \in \mathcal{S}^*$, which are given by

$$\frac{\partial Y_k^*}{\partial \beta_i} = Y_k^* \left(\frac{\partial \mu_k^*}{\partial \beta_i} V^*(k) + \frac{\partial V^*(k)}{\partial \beta_i} \mu_k^* \right), \forall k \in \mathcal{S}^*, \tag{23}$$

and

$$\frac{\partial V^*}{\partial \beta_i} = (I - H)^{-1} (L^i e + h), \tag{24}$$

where β_i is a parameter, L^i, H are two matrices of size $|\mathcal{S}^*| \times |\mathcal{S}^*|$ and h is a vector of size $|\mathcal{S}^*|$ with entries

$$L_{ka}^i = \frac{1}{\mu_k^*} \frac{\partial M_{ka}^*}{\partial \beta_i} \frac{(Y_a^*)^{\phi_{ka}^*}}{(Y_k^*)} + \frac{1}{\mu_k^*} M_{ka}^* \ln(Y_a^*) \frac{(Y_a^*)^{\phi_{ka}^*}}{(Y_k^*)} \frac{\partial \phi_{ka}^*}{\partial \beta_i} + M_{ka}^* \ln(Y_a^*) \frac{(Y_a^*)^{\phi_{ka}^*}}{(Y_k^*)} \frac{\partial \mu_a^*}{(\mu_a^*)^2 \partial \beta_i},$$

and

$$H_{ka} = M_{ka}^* \frac{(Y_a^*)^{\phi_{ka}^*}}{(Y_k^*)} \text{ and } h_k = -\frac{\partial \mu_k^*}{(\mu^*)_k^2 \partial \beta_i} \ln(Y_k^*) \text{ and } \phi_{ka}^* = \mu_k^* / \mu_a^*.$$

Mai et al. (2015c) suggest that deriving the gradients based on V^* is better than based on Y^* for numerical reasons. The system of linear equations in (24) can be solved quickly for large scale-systems.

5. Recursive cross-nested logit model

The cross-nested logit (CNL) model is an instance of the network MEV model that allows each alternative to belong several different nests. It has been mentioned for the first time by Vovsha (1997) in the context of a mode choice survey in Israel. Papola (2004) has shown that a specific CNL model can be derived for any given homoscedastic variance-covariance matrix. Fosgerau et al. (2013b) show that any additive random utility model can be approximated by the cross-nested logit model. Therefore, the CNL model, with closed forms for the choice probabilities, becomes a serious competitor for the probit model.

In this section we present the RCNL model, which is an instance of the RNMEV model where the choice at each state is a CNL model based on the formulation proposed by Ben-Akiva et al. (1999). In this setting, the CNL model at state k is a network MEV model given by a network of correlation structure $\mathcal{G}^k$ where the corresponding CPGF $G_k(y)$ is

$$G_k(y) = \sum_{m \in S_k(r)} \left(\sum_{a \in S_k(m)} \alpha_{ma}^k y_a^{\xi_m^k} \right)^{\xi_r^k / \xi_m^k}, \tag{25}$$

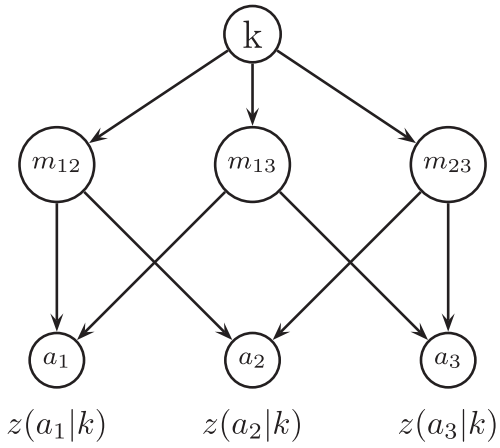

Fig. 4. A cross-nested structure at link k.

(recall that r is the root of the network MEV model at state k). We remark that $S_k(r)$ is also the set of nests. For each state $a \in S(k)$ we denote $z(a|k) = v(a|k) + V(a)$ and note that a next state is chosen by maximizing the sum of $z(a|k)$ and the random term $\epsilon(a|k)$, $\forall a \in S(k)$. In order to model the correlation between the successor states $a \in S(k)$, we use the paired combinatorial logit proposed by Koppelman and Wen (2000), i.e. a CNL model where each pair of states belongs to only one nest, and each nest contains only one pair of states. The number of pairs in the set $S(k)$ is $\frac{1}{2}|S(k)|.(|S(k)|-1)$, so it is also the number of nests at choice stage k. Fig. 4 illustrates an example where there are three successor states from k. Accordingly, there are three nests m_{12}, m_{13} and m_{23}. Two states a_i and a_j belong to nest m_{ij}, $\forall i, j = 1, 2, 3$.

Given this correlation structure, based on Papola (2004), the correlation between two given states a_i, $a_j \in S(k)$ can be approximated as

$$\widehat{\text{Corr}}\big(z(a_i|k), z(a_j|k)\big) = \sum_{m \in S_k(r)} (\alpha^k_{ma_i}\alpha^k_{ma_j})^{0.5\xi^k_r/\xi^k_m}\left(1 - \left(\frac{\xi^k_r}{\xi^k_m}\right)^2\right). \tag{26}$$

Since there is only one nest m_{ij} that both a_i and a_j belong to, so

$$\widehat{\text{Corr}}\big(z(a_i|k), z(a_j|k)\big) = (\alpha^k_{m_{ij}a_i}\alpha^k_{m_{ij}a_j})^{0.5\xi^k_r/\xi^k_{m_{ij}}}\left(1 - \left(\frac{\xi^k_r}{\xi^k_{m_{ij}}}\right)^2\right). \tag{27}$$

Hence, the correlation between a_i, $a_j \in S(k)$ can be modeled using the parameters $\alpha^k_{m_{ij}a_i}$, $\alpha^k_{m_{ij}a_j}$, ξ^k_r and ξ^k_m. It is important to note that if the choice model at state k is MNL, then for any two states a_i, $a_j \in S(k)$, $a_i \neq a_j$, we have $\text{Corr}\big(z(a_i|k), z(a_j|k)\big) = 0$, and the variance-covariance matrix is diagonal. So, the RCNL model allows to exhibit a more general correlation structure at each choice stage, compared to the NRL and RL models.

Abbé et al. (2007) note that Papola's approximation in (27) can overestimate the correlation in some cases and bias the choice probabilities provided by the CNL model. However, they also comment that these biases do not seem to be large in their examples.

For the estimation, the RCNL model is a member of the RNMEV model, so this can be estimated by applying the method presented in Section 3, i.e., estimating the NRL model using the integrated network $\mathcal{G}^*$ and using Theorems 1 and 2. According to Papola (2004), a normalization required for the CNL model given by (25) is

$$\sum_{m \in S_k(r)} \left(\alpha^k_{ma}\right)^{\xi^k_r/\xi^k_m} = 1. \tag{28}$$

Finally, the integrated network has larger state space, compared to the original (i.e. $\mathcal{G}$). Based on Proposition 1, the numbers of states and arcs in the integrated network are

$$|\mathcal{S}^*| = |\tilde{\mathcal{S}}| + \frac{1}{2}\sum_{k \in \mathcal{S}} (|S(k)| - 1)|S(k)|,$$

and

$$|\mathcal{A}^*| = \frac{3}{2} \times \sum_{k \in \mathcal{S}} (|S(k)| - 1)|S(k)|.$$

6. Application to real data

We illustrate the proposed methods and models by using the same data used in Fosgerau et al. (2013a); Mai et al. (2015c), which has been collected in Borlänge, Sweden. The network is composed of 3077 nodes and 7459 links and is uncongested so travel times can be assumed static and deterministic. The observations consist of 1832 trips corresponding to simple paths with a minimum of five links. Moreover, there are 466 destinations, 1420 different origin-destination (OD) pairs and more than 37,000 link choices in this sample. We present estimation and prediction results for the RCNL model. For the sake of comparison we include the results from the NRL model (Mai et al., 2015c). For the other recursive model, i.e., the RL model, reader can consult Mai et al. (2015c) for details.

6.1. Model specification

We use the same attributes as Mai et al. (2015c) for the instantaneous utilities. We note that the other recursive route choice models in Mai et al. (2015c) and Fosgerau et al. (2013a) define the models and utilities based on links in the network to capture turn attributes. Since the models presented in this paper are based on a network of states, we then define the utility specifications based on states and note that each state refers to a link in the real transport network.

Five attributes are considered: travel time $TT(a)$ of action a, a left turn dummy $LT(a|k)$ that equals one if the turn angle from k to a is larger than 40° and less than 177° , a u-turn dummy $UT(a|k)$ that equals one if the turn angle is larger than 177, and a state constant $LC(a)$ that equals one except the dummy d which equals zero. The rationale behind using $LC(a)$ in the instantaneous utilities is to penalize paths with many crossings. The fifth attribute is $LS(a)$ (for a detailed description see Fosgerau et al., 2013a) and it has been computed using a linear in parameters formulation of the aforementioned four attributes using parameters $\tilde{\beta}_{TT} = -2.5,\ \tilde{\beta}_{LT} = -1,\ \tilde{\beta}_{LC} = 0.4,\ \tilde{\beta}_{UT} = -4$.

We specify the deterministic utilities for different model specifications with respect to state a given state k, $k \in \mathcal{S}, a \in S(k)$

$$v^{\text{NRL}}(a|k;\beta) = v^{\text{RCNL}}(a|k;\beta) = \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k),$$
$$v^{\text{NRL-LS}}(a|k;\beta) = v^{\text{RCNL-LS}}(a|k;\beta) = \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) + \beta_{LS}LS(a).$$

In the NRL model in Mai et al. (2015c), the scale of the random terms are defined as exponential functions of the model parameters. Mai et al. (2015c) use the travel time, link size and number of successor states to define the scale as

$$\mu_k^{NRL}(\omega) = e^{\omega_{TT}TT(k)+\omega_{LS}LS(k)+\omega_{OL}OL(k)},\ \forall k \in \tilde{\mathcal{S}},$$

where $OL(k)$ is the number of successor states from k i.e. $OL(k) = |S(k)|$ (it is also the number of outgoing links from the sink node of a link in the road network). Note that this NRL is based on network $\mathcal{G}$ and differs from the NRL model based on the integrated network $\mathcal{G}^*$. The latter is used for the estimation of the RCNL model.

The CNL model at each state includes several structure parameters. Our network contains more than 7000 states/links, leading to more than 7000 CNL models. So it is not possible to estimate all the parameters. Similar to Mai et al. (2015c) we define parameters ξ_r^k and ξ_m^k, $\forall m \in S_k(r)$ as exponential functions of the respective state attributes. More precisely, since at choice stage k, the root r of the network MEV model is identical to state k, so ξ_r^k is defined as

$$\xi_r^k(\omega) = e^{\omega_{TT}TT(k)+\omega_{LS}LS(k)+\omega_{OL}OL(k)},\ \forall k \in \mathcal{S}. \tag{29}$$

Indeed, by definition $\xi_r^k(\omega) > 0$, $\forall \omega$. Moreover, Eq. (27) suggests that the correlation between two given states a_i, a_j depends on the ratio $\frac{\xi_{m_{ij}}^k}{\xi_r^k}$. We therefore define this fraction as an exponential function of the attributes associated with states a_i and a_j as

$$\frac{\xi_{m_{ij}}^k}{\xi_r^k} = e^{\lambda_{TT}\left(TT(a_i)+TT(a_j)\right)+\lambda_{LS}\left(LS(a_i)+LS(a_j)\right)+\lambda_{OL}\left(OL(a_i)+OL(a_j)\right)}, \tag{30}$$

or equivalently the scale $\xi_{m_{ij}}^k$ associated with nest m_{ij} is defined as

$$\xi_{m_{ij}}^k = \xi_r^k e^{\lambda_{TT}\left(TT(a_i)+TT(a_j)\right)+\lambda_{LS}\left(LS(a_i)+LS(a_j)\right)+\lambda_{OL}\left(OL(a_i)+OL(a_j)\right)}, \tag{31}$$

where ξ_r^k is defined in (29). The CNL model requires constraints on the scale parameters which are $\xi_{m_{ij}}^k \geq \xi_r^k$, $\forall m_{ij} \in S_k(r)$. We therefore impose these constraints by restricting the parameter λ to being positive, i.e., $\lambda_{TT}, \lambda_{LS}, \lambda_{OL} \geq 0$.

In the CNL model at state k, the value $(\alpha_{ma}^k)^{\xi_r^k/\xi_m^k}$ reflects the level of membership of alternative a to nest m (Abbé et al., 2007). Indeed, it is impossible to estimate all the parameters α in the network. We therefore assume that each state $a \in S(k)$, the levels of a to all the nests m such that $a \in S_k(m)$ are equal. Based on the normalization in (28) and as the number of nests that each node $a \in S(k)$ belong to is $|S(k)| - 1$, the parameters are specified as

$$\alpha_{ma}^k = \left(\frac{1}{|S(k)|-1}\right)^{\xi_m^k/\xi_r^k}. \tag{32}$$

Hence, the correlation between two states a_i, $a_j \in S(k)$ can be approximated as

$$\widehat{\mathrm{Corr}}\big(z(a_i|k), z(a_j|k)\big) = \frac{1}{|S(k)-1|}\left(1-\left(\frac{\xi_r^k}{\xi_{m_{ij}}^k}\right)^2\right),$$

where $\frac{\xi_r^k}{\xi_{m_{ij}}^k}$ is defined in (30). In summary, the instantaneous utilities are

$$\begin{aligned}
u^{\mathrm{NRL}}(a|k;\beta,\omega) &= v^{\mathrm{NRL}}(a|k;\beta) + \frac{1}{\mu_k^{NRL}(\omega)}(\epsilon(a)-\gamma),\\
u^{\mathrm{NRL\text{-}LS}}(a|k;\beta,\omega) &= v^{\mathrm{NRL\text{-}LS}}(a|k;\beta) + \frac{1}{\mu_k^{NRL}(\omega)}(\epsilon(a)-\gamma),\\
u^{\mathrm{RCNL}}(a|k;\beta,\omega,\lambda) &= v^{\mathrm{RCNL}}(a|k;\beta) + \epsilon(a|k;\omega,\lambda),\\
u^{\mathrm{RCNL\text{-}LS}}(a|k;\beta,\omega,\lambda) &= v^{\mathrm{RCNL\text{-}LS}}(a|k;\beta) + \epsilon(a|k;\omega,\lambda),
\end{aligned}$$

where $\epsilon(a)$ is i.i.d extreme value type I and $\epsilon(a|k;\omega)$, $a \in S(k)$, have an MEV distribution with the CPGF $G_k(y)$ specified in (25) and the structure parameters specified in (29), (31) and (32). Indeed, if $\omega = 0$ then the NRL model becomes the RL model and if $\lambda = 0$ then $\xi_r^k = \xi_m^k$, $\forall k \in \mathcal{S}$, $m \in S_k(r)$, so the G_k function in (25) becomes

$$G_k(y) = \sum_{a\in S(k)} y_a^{\xi_r^k} = \sum_{a\in S(k)} y_a^{\mu_k},$$

meaning that the choice model at state k is MNL, and the RCNL model becomes exactly the NRL model. Finally, the maximum likelihood estimation of the RCNL model is a constrained optimization problem and can be expressed as

$$\max_{\substack{\beta,\omega,\lambda\\ \lambda\geq 0}} LL^{RCNL}(\beta,\omega,\lambda).$$

6.2. Estimation results

We report the estimation results in Table 1 for the four specifications NRL, NRL-LS, RCNL and RCNL-LS. The results are comparable to those previously published using the same data. The β estimates have their expected signs and are highly significant. For both the NRL and RCNL models, the ω estimates are negative for travel time and positive for left turns and link constant. All the ω estimates for the RCNL models are significantly different from zero. However, for the NRL model, $\hat{\omega}_{TT}$ is not significantly different from zero when the LS attribute is included in the instantaneous utilities.

We now turn our attention to the λ estimates. The λ estimates are very close to zero for travel time and link size but significantly different from zero for the parameters associated with the number of successor states. Note that we do not provide standard errors and t-tests for the estimates that are on the bound (close to 0) since the respective gradient component values are not close to zero. The λ estimates indicate that only the attribute *OL* affects the correlation between successor states.

The final LL values are reported in Table 1 and we also report the likelihood ratio test in Table 2. Similar to Fosgerau et al. (2013a) and Mai et al. (2015c), we observe a significant improvement in final LL values when we include the LS attribute to the instantaneous utilities. The RCNL models have better fit than the NRL models and the best model in term of in-sample fit is the RCNL-LS. Moreover, the final LL function given by the RCNL without the LS is larger than the one given by the NRL model with the LS attribute.

We note that the estimation of the RCNL model requires estimating the NRL model on the integrated network. The real network has 7288 states (i.e. links) and the integrated network has 31,373 states, so there are 24,084 new states added to the original network. The number of arcs in the integrated network is 72,252, compared to 20,202 arcs in the original network. So the estimation of the RCNL model is more expensive than the NRL and RL models. Solving the value functions in the RCNL model needs from 300 to 700 iterations to converge to the fixed point solutions while the NRL needs less than 300 iterations. Furthermore, using the dynamic accuracy and switching approaches, we are able to double the speed of the value iteration method.

6.3. Prediction results

In this section we compare the prediction performance of the different models. Similar to Mai et al. (2015c), we use a cross validation approach, i.e., the sample of observations is divided into two sets by drawing observations uniformly: one set contains 80% of the observations is used for estimation, and the other set (20% of the observations) is used as holdout to evaluate the predicted probabilities by applying the estimated model. We generate 20 holdout samples of the same size by reshuffling the real sample and we use the LL loss to evaluate the prediction performance.

Table 1
Estimation results.

Parameters	NRL	NRL-LS	RCNL	RCN-LS
$\widehat{\beta}_{TT}$	−1.854	−2.139	−1.378	−1.567
Rob. Std. Err.	0.132	0.145	0.080	0.077
Rob. t-test(0)	−14.05	−14.75	−17.296	−20.336
$\widehat{\beta}_{LT}$	−0.679	−0.748	−0.517	−0.568
Rob. Std. Err.	0.043	0.047	0.018	0.019
Rob. t-test(0)	−15.79	−15.91	−29.528	−30.105
$\widehat{\beta}_{LC}$	−0.258	−0.224	−0.065	−0.072
Rob. Std. Err.	0.016	0.015	0.013	0.011
Rob. t-test(0)	−16.13	−14.93	−4.865	−6.484
$\widehat{\beta}_{UT}$	−3.340	−3.301	−2.907	−2.964
Rob. Std. Err.	0.200	0.207	0.094	0.099
Rob. t-test(0)	−16.7	−15.95	−30.963	−30.057
$\widehat{\beta}_{LS}$	–	−0.155	-	−0.115
Rob. Std. Err.	-	0.013	-	0.007
Rob. t-test(0)	-	−11.92	-	−16.140
$\widehat{\omega}_{TT}$	−0.515	−0.341	−0.637	−0.443
Rob. Std. Err.	0.255	0.288	0.220	0.216
Rob. t-test(0)	−2.02	−1.18	−2.898	−2.053
$\widehat{\omega}_{LS}$	0.674	0.581	0.192	0.157
Rob. Std. Err.	0.093	0.09	0.031	0.025
Rob. t-test(0)	7.25	6.46	6.224	6.275
$\widehat{\omega}_{OL}$	0.086	0.092	0.027	0.021
Rob. Std. Err.	0.015	0.016	0.003	0.002
Rob. t-test(0)	5.73	5.75	9.846	9.167
$\widehat{\lambda}_{TT}$	–	–	2.63E-04	1.84E-08
Rob. Std. Err.	–	–	–	–
Rob. t-test(0)	–	–	–	–
$\widehat{\lambda}_{LS}$	–	–	2.85E-05	1.74E-07
Rob. Std. Err.	–	–	–	–
Rob. t-test(0)	–	–	–	–
$\widehat{\lambda}_{OL}$	–	–	0.475	0.483
Rob. Std. Err.	–	–	0.012	0.012
Rob. t-test(0)	–	–	41.151	41.230
$LL(\widehat{\beta})$	−6187.9	−5952	−5885.5	−5675.4

Table 2
Likelihood ratio test results.

Models	χ^2	p-value
NRL & NRL-LS	471.8	1.30e-104
NRL & RCNL	604.8	9.18e-131
NRL-LS & RCNL-LS	553.2	1.41e-119
RCNL & RCNL-LS	420.2	2.21e-93

For each holdout sample i, $0 \leq i \leq 20$, we estimate the parameters $\hat{\beta}_i$ of the corresponding training sample PS_i and this vector of parameters is used to compute the test errors err_i

$$err_i = -\frac{1}{|PS_i|} \sum_{\sigma_j \in PS_i} \ln P(\sigma_j, \hat{\beta}_i),$$

where $|PS_i|$ is the size of prediction sample PS_i. In order to have unconditional test error values, we compute the average of err_i over samples as follows

$$\overline{err}_p = \frac{1}{p} \sum_{i=1}^{p} err_i, \quad \forall 1 \leq p \leq 20. \tag{33}$$

The values of $\overline{err}_p$, $1 \leq p \leq 20$, are plotted in Fig. 5 and Table 3 reports the average of the test error values over 20 samples given by the NRL, NRL-LS, RCNL, RCNL-LS models. The lower test error values the better the model in prediction.

The prediction results show that the models with the LS attribute perform better than those without. The RCNL models have better prediction performances than the NRL models. The RCNL-LS performs the best in fit and prediction among the considered models.

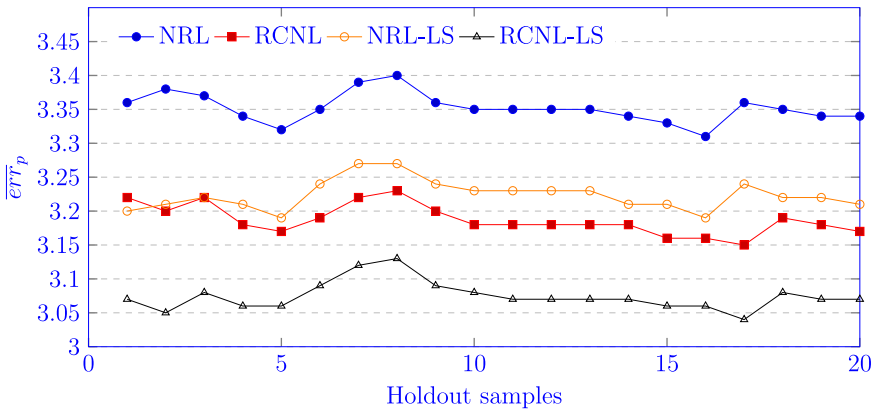

Fig. 5. Average of the test error values over holdout samples.

Table 3
Average of test error values over 20 holdout samples.

NRL	NRL-LS	RCNL	RCNL-LS
3.34	3.21	3.17	3.07

7. Discussion on the model specifications and applications

In this paper, the RCNL model is specified so that the covariance at each choice state can be computed and estimated, under the assumption that the correlation between two outgoing states could be affected by their characteristics. We however note that the correlation in recursive models can be captured in other different ways. For examples, the correlation between different state alternatives could be driven by the topology of the corresponding subsequent networks, or could be designed specifically for other types of networks (e.g. dynamic, multi-modal networks). The method proposed in this paper provide a way to deal with different designs for the correlation, and could have an impact in other route choice applications.

It is important to note that the RNMEV model proposed in this paper and other variants (Fosgerau et al., 2013a; Mai et al., 2015a; 2015b; 2015c) may be useful for traffic simulation. In this context, one needs to compute the next-state choice probabilities $P(a|k)$, $k \in \mathcal{S}, a \in S(k)$. These probabilities are computed using the value functions, and the link flows can be obtained by inverting matrix $(I - P^T)$, where I is the identity matrix. Baillon and Cominetti (2008) show that $(I - P^T)$ is invertible. In this paper we show that solving the value functions in the RNMEV model could be greatly simplified by adding artificial states to the transport network. This method is therefore useful for prediction and simulation, in addition to the maximum likelihood estimation.

Moreover, in some applications, the correlation structures at choice states may be observation specific, and the corresponding model could become more expensive for both estimation and prediction. In this context, it is important to accelerate the process of creating the artificial networks. The observations can also be aggregated into different groups such that the correlation structures are group specific, then we can just create a new network once for each group and reduce the computational cost.

We recall that the recursive models are developed based on the dynamic discrete choice framework proposed by Rust (1987). In this paper we show that network MEV models can be integrated into dynamic discrete choice models to flexible capture the correlation in route choice applications. In general, this idea can be adapted and applied to other dynamic discrete choice applications, as relaxing the i.i.d. assumption in Rust's model is a topic of interest. Moreover, the work reported in this paper is also related to another context than route choice where there is an issue associated with estimating MEV models with large choice sets (Mai et al., 2015e).

8. Conclusion

This paper has considered the RNMEV model, a generalized recursive route choice model, where the choice model at each stage can be any member of the network MEV model. We have introduced a method of integrating correlation structures for simplifying the contraction mapping given by the Bellman's equation of the RNMEV model. This approach allowed us to use the contraction mapping from the NRL model and the methods proposed in Mai et al. (2015c) to quickly estimate the RNMEV model on real data.

We have applied the proposed method to the estimation of the RCNL model, a member of the RNMEV model where the model at each choice stage is a cross-nested logit model. We have showed that the RCNL can exhibit a more general

correlation structure, compared to the RL and NRL models. We have provided numerical results using a real data. The RCNL model can be estimated quickly using the new method. Moreover, the parameter estimates are sensible, and the RCNL model has significantly better fit than the NRL model. We have also provided a cross-validation study suggesting that RCNL and RCNL-LS are better than the NRL and NRL-LS models.

In this paper, we have used an unimodal network and observations of trips made by car but the model is not restricted to this type of network. In the future research, we plan to apply the RNMEV to different types of networks where the correlation at each choice stage need to be taken into account, e.g., dynamic networks (state is time and location), and multi-modal networks (state is location and mode). Other designs for correlation, e.g. based on the topology of the network, would also be interesting to investigate.

Finally, we note that the method of integrating correlation structures is not restricted to route choice applications. It can be used with the nested fixed point algorithm (Rust, 1987) to estimate dynamic discrete choice models with MEV random terms.

Acknowledgments

This research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC), discovery grant 435678-2013 held by Emma Frejinger. I am grateful for valuable comments and suggestions from Fabian Bastin, Mogens Fosgerau and Emma Frejinger. The paper has also been improved thanks to the comments of the anonymous reviewer.

Appendix A. Proof of Theorem 1

Consider the network of correlation structure $\mathcal{G}_k = (\mathcal{S}_k, \mathcal{A}_k, C_k)$ at state k. In order to prove the result we derive the function $G_k\left(e^{v(a|k)}(Y_a^*)^{1/\mu_a}, a \in S(k)\right)$ using (6) and (7), and note that $G_k(y) = G_k^r(y)$, where r is the root of $\mathcal{G}_k$. For notational simplicity we denote y_k^* as a vector of size $|S(k)|$ with entries $e^{v(a|k)}(Y_a^*)^{1/\mu_a}$, for all $a \in S(k)$. We also note that, as discussed in Section 3, state k is also the root r and the choice set C_k is identical to $S(k)$.

We first introduce some definitions. For each state $i \in \mathcal{S}_k$, we denote $\mathcal{L}^k(i)$ as the length (defined as number of arcs) of the longest sequences of states (or paths) connecting i and all $j \in S(k)$ via states in $\mathcal{S}_k$. $\mathcal{L}^k(i)$ is finite since the network $\mathcal{G}_k$ is cycle-free. For any integer number $p \geq 0$ we denote $\mathcal{T}^k(p)$ as the set of state i such that $i \in \mathcal{S}_k$ and $\mathcal{L}^k(i) = p$. In other words

$$\mathcal{T}^k(p) = \{i | i \in \mathcal{S}_k, \mathcal{L}^k(i) = p\}, \ \forall p \in \mathbb{N}.$$

We have the following proposition, which is easy to verify

Proposition 2.

(i) $\mathcal{T}^k(0) = S(k)$, $\bigcup_{t=1}^{\mathcal{L}^k(r)} = \mathcal{S}_k$.
(ii) $\mathcal{T}^k(p) \cap \mathcal{T}^k(q) = \emptyset$, $p, q \geq 0$, $p \neq q$.
(iii) *Given state* $i \in \mathcal{T}^k(p)$, $p \geq 1$, *if* $j \in S^*(i)$ *then* $j \in \bigcup_{t=0}^{p-1} \mathcal{T}^k(t)$.

Proof. (i) and (ii) are trivial to verify. For (iii), we suppose that $j \notin \bigcup_{t=0}^{p-1} \mathcal{T}^k(t)$, then there exists a number $p' \geq p$ such that $j \in \mathcal{T}^k(p')$. It means that there exists a sequence of length p' connecting j and states in $S(k)$. Moreover, since $j \in S^*(i)$ and from the fact that $i \notin S(k)$ (because $i \in \mathcal{T}^k(p)$ and $p \geq 1$) we have $j \in S_k(i)$ due to Proposition 1(iv). Consequently, there exists a sequence of length $p' + 1 > p$ connecting i and states in $S(k)$. This is in contradiction with the assumption that $i \in \mathcal{T}^k(p)$. So j has to be in $\bigcup_{t=0}^{p-1} \mathcal{T}^k(t)$ and (iii) is proved. □

For all $i \in \mathcal{S}_k$, the values of $G_k^i(y_k^*)$ can be computed based on (6) and (7) as

$$G_k^i(y_k^*) = \left(e^{v(i|k)}(Y_i^*)^{1/\mu_i}\right)^{\xi_i^k}, \ \forall i \in S(k). \tag{34}$$

For each $i \in \mathcal{S}_k \backslash S(k)$ we have

$$G_k^i(y_k^*) = \sum_{j \in S_k(i)} \alpha_{ij}^k G_k^j(y_k^*)^{\xi_i^k/\xi_j^k}. \tag{35}$$

We introduce the following lemma

Lemma 1. *Given state* $k \in \mathcal{S}$, *if* $G_k^i(y_k^*)$, $\forall i \in \mathcal{S}_k$, *are computed based on (34) and (35) then*

$$G_k^i(y_k^*) = Y_i^*, \ \forall i \in \mathcal{T}^k(p), \forall p \in \mathbb{Z}^+. \tag{36}$$

Proof. Based on the definitions in (10) and (11), (34) can be written equivalently as

$$G_k^i(y_k^*) = e^{\xi_i^k v(i|k)}(Y_i^*)^{\xi_i^k/\mu_i^*}, \ \forall i \in S(k). \tag{37}$$

Here we remark that $\xi_i^k \neq \mu_i^*$, $\forall i \in S(k)$, and $\xi_i^k = \mu_i^*$ $\forall i \in \mathcal{S}_k \backslash S(k)$. Due to Proposition 1(iv) we have $S_k(i) = S^*(i)$, $\forall i \in \mathcal{S}_k \backslash S(k)$, so (35) can be written as

$$\begin{aligned} G_k^i(y_k^*) &= \sum_{j \in S^*(i)} \alpha_{ij}^k G_k^j(y_k^*)^{\xi_i^k/\xi_j^k} \\ &= \sum_{\substack{j \in S^*(i) \\ j \notin S(k)}} e^{\mu_i^* v^*(j|i)} \left(G_k^j(y_k^*)\right)^{\mu_i^*/\mu_j^*} + \sum_{\substack{j \in S^*(i) \\ j \in S(k)}} e^{\mu_i^*(v^*(j|i) - v(j|k))} \left(G_k^j(y_k^*)\right)^{\mu_i^*/\xi_j^k}. \\ &\forall i \in \mathcal{S}_k \backslash S(k). \end{aligned} \tag{38}$$

Substituting (37) into (38) we obtain

$$G_k^i(y_k^*) = \sum_{\substack{j \in S^*(i) \\ j \notin S(k)}} e^{\mu_i^* v^*(j|i)} \left(G_k^j(y_k^*)\right)^{\mu_i^*/\mu_j^*} + \sum_{\substack{j \in S^*(i) \\ j \in S(k)}} e^{\mu_i^* v^*(j|i)} (Y_j^*)^{\mu_i^*/\mu_j^*}, \ \forall i \in \mathcal{S}_k \backslash S(k). \tag{39}$$

Now we prove the result by induction. For $p = 1$, according to Proposition 2(i)–(iii) we have the fact that for each $i \in \mathcal{T}^k(1)$, if $j \in S^*(i)$ then $j \in \mathcal{T}^k(0)$, or equivalently $j \in S(k)$. Thus, (39) can be written as

$$G_k^i(y_k^*) = \sum_{j \in S^*(i)} e^{\mu_i^* v^*(j|i)} (Y_j^*)^{\mu_i^*/\mu_j^*}, \ \forall i \in \mathcal{T}^k(1). \tag{40}$$

So from (15) and (40) we have $G_k^i(y_k^*) = Y_i^*$ $\forall i \in \mathcal{T}^k(1)$, meaning that (36) is true for $p = 1$. Now we assume that the result is true for $p \geq 1$. In other words

$$G_k^i(y_k^*) = Y_i^*, \ \forall i \in \bigcup_{t=1}^{p} \mathcal{T}^k(t).$$

We will prove that the result is also true for $p+1$. For each state $i \in \mathcal{T}^k(p+1)$, according to Proposition 2(iv), if $j \in S^*(i) \backslash S(k)$ then $j \in \bigcup_{t=1}^{p} \mathcal{T}^k(t)$. Consequently, by assumption, $G_k^j(y_k^*) = Y_j^*$ $\forall j \in S^*(i) \backslash S(k)$. Hence, (39) can be written as

$$\begin{aligned} G_k^i(y_k^*) &= \sum_{\substack{j \in S^*(i) \\ j \notin S(k)}} e^{\mu_i^* v^*(j|i)} \left(Y_j^*\right)^{\mu_i^*/\mu_j^*} + \sum_{\substack{j \in S^*(i) \\ j \in S(k)}} e^{\mu_i^* v^*(j|i)} (Y_j^*)^{\mu_i^*/\mu_j^*} \\ &= \sum_{j \in S^*(i)} e^{\mu_i^* v^*(j|i)} \left(Y_j^*\right)^{\mu_i^*/\mu_j^*}, \ \forall i \in \mathcal{T}^k(p+1), \end{aligned} \tag{41}$$

so $G_k^i(y_k^*) = Y_i^*$, $\forall i \in \mathcal{T}^k(p+1)$, because of (15). This validates (36) for $p+1$, as required. □

We note that if $p = \mathcal{L}^k(k)$ then $G_k^k(y_k^*) = Y_k^*$, or $Y_k^* = G_k(y_k^*)$. Hence, Theorem 1 is proved.

Appendix B. Proof of Theorem 2

We consider the network $\mathcal{G}_k = (\mathcal{S}_k, \mathcal{A}_k, C_k)$ at state $k \in \mathcal{S}$. Under the hypotheses of Theorem 2 we have the fact that $y_k = y_k^*$ (we recall that y_k is a vector of size $|S(k)|$ with elements $e^{v(a|k)}(Y_a)^{1/\mu_a}$, $\forall a \in S(k)$). So from (8), the choice probability $P(a|k)$ $\forall a \in S(k)$ given by the network MEV model at state k is

$$P(a|k) = \sum_{[j_0, \ldots, j_l] \in \Omega^k(a)} \prod_{t=0}^{l-1} \frac{\alpha_{j_t j_{t+1}}^k (G_k^{j_{t+1}}(y_k^*))^{\xi_{j_t}^k/\xi_{j_{t+1}}^k}}{G_k^{j_t}(y_k^*)}. \tag{42}$$

Given two states $i, j \in \mathcal{S}_k$, $j \in S^*(i)$, we consider two cases: $j \notin S(k)$ or $j \in S(k)$.

- If $j \notin S(k)$, then according to Lemma 1, we have $G_k^i(y_k^*) = Y_i^*$ and $G_k^j(y_k^*) = Y_j^*$. Furthermore, from the definitions in (10), (11) and Eq. (16) we obtain

$$\frac{\alpha_{ij}^k (G_k^j(y_k^*))^{\xi_i^k/\xi_j^k}}{G_k^i(y_k^*)} = e^{\mu_i^* v^*(j|i)} \frac{(Y_j^*)^{\mu_i^*/\mu_j^*}}{Y_i^*} = P^*(j|i). \tag{43}$$

- If $j \in S(k)$, then from (37) we have

$$\begin{aligned} \frac{\alpha_{ij}^k (G_k^j(y_k^*))^{\xi_i^k/\xi_j^k}}{G_k^i(y_k^*)} &= \frac{e^{\mu_i^*(v^*(j|i) - v(j|k))}}{Y_i^*} e^{\mu_i^* v(j|k)} (Y_j^*)^{\mu_i^*/\mu_j^*} \\ &= e^{\mu_i^* v^*(j|i)} \frac{(Y_j^*)^{\mu_i^*/\mu_j^*}}{Y_i^*} = P^*(j|i). \end{aligned} \tag{44}$$

So from (43) and (44) we obtain

$$\frac{\alpha_{ij}^k (G_k^j(y_k^*))^{\xi_i^k/\xi_j^k}}{G_k^i(y_k^*)} = P^*(j|i),\ \forall i \in \mathcal{S}_k \backslash S(k),\ j \in S^*(i),$$

and the choice probability $P(a|k)$ in (42) can be computed as

$$P(a|k) = \sum_{[j_0,\dots,j_I] \in \Omega^k(a)} \prod_{t=0}^{I-1} P^*(j_{t+1}|j_t).$$

References

Abbé, E., Bierlaire, M., Toledo, T., 2007. Normalization and correlation of cross-nested logit models. Transp. Res. Part B 41 (7), 795–808.
Baillon, J.-B., Cominetti, R., 2008. Markovian traffic equilibrium. Math. Program. 111 (1–2), 33–56.
Ben-Akiva, M., 1973. The structure of travel demand models. MIT Phd thesis.
Ben-Akiva, M., Bierlaire, M., 1999. In: Hall, R. (Ed.), Discrete choice methods and their applications to short-term travel decisions. Kluwer, pp. 5–34.
Daly, A., Bierlaire, M., 2006. A general and operational representation of generalised extreme value models. Transp. Res. Part B 40 (4), 285–305.
Fosgerau, M., Frejinger, E., Karlström, A., 2013a. A link based network route choice model with unrestricted choice set. Transp. Res. Part B 56, 70–80.
Fosgerau, M., McFadden, M., Bierlaire, M., 2013b. Choice probability generating functions. J. Choice Modell. 8, 1–18.
Koppelman, F., Wen, C.-H., 2000. The paired combinatorial logit model: properties, estimation and application. Transp. Res. Part B 34, 75–89.
Mai, T., Bastin, F., Frejinger, E., 2015a. A decomposition method for estimating recursive logit based route choice models, CIRRELT-2015-66. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation.
Mai, T., Bastin, F., Frejinger, E., 2015b. Comparing regret minimization and utility maximization for route choice using the recursive logit model. CIRRELT-2015-38. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation.
Mai, T., Fosgerau, M., Frejinger, E., 2015c. A nested recursive logit model for route choice analysis. Transp. Res. Part B 75 (1), 100–112.
Mai, T., Frejinger, E., Bastin, F., 2015. A misspecification test for logit based route choice models. Econ. Transp. 4 (4), 215–226.
Mai, T., Frejinger, E., Fosgerau, M., Bastin, F., 2015e. A dynamic programming approach for quickly estimating large-scale MEV models. CIRRELT-2015-24. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation.
McFadden, D., 1978. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Modelling the choice of residential location. North-Holland, Amsterdam, pp. 75–96.
Papola, A., 2004. Some developments on the cross-nested logit model. Transp. Res. Part B 38 (9), 833–851.
Rust, J., 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. Econometrica 55 (5), 999–1033.
Small, K.A., 1987. A discrete choice model for ordered alternatives. Econometrica 55 (2), 409–424.
Swait, J., 2001. Choice set generation within the generalized extreme value family of discrete choice models. Transp. Res. Part B 35, 643–666.
Vovsha, P., 1997. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. Transp. Res. Rec. 1607, 6–15.
Vovsha, P., Bekhor, S., 1998. Link-nested logit model of route choice Overcoming route overlapping problem. Transp. Res. Rec. 1645, 133–142.
Wen, C.-H., Koppelman, F., 2001. The generalized nested logit model. Transp. Res. Part B 35 (7), 627–641.