Recursive network MEV model for route choice analysis

Tien Mai *

July 31, 2015

Abstract

This paper presents a general and operational representation of the recursive models for route choice analysis. We extend the nested recursive logit model (NRL) (Mai et al., 2015) by allowing the choice at each stage to be a network multivariate extreme value (MEV) model (Daly and Bierlaire, 2006) instead of the multinomial logit (MNL) model. Similar to the NRL model, the choice of path is modeled as a sequence of state choices and the model does not require any sampling of choice sets. Furthermore, the model can be consistently estimated and efficiently used for prediction

The main challenge is on the computation of the value functions which are solutions to a complex non-linear system. We present a novel approach where a new network is created by integrating the networks of correlation structures given by the network MEV models into the transport network. We show similarities between the RNMEV model and the NRL model on the integrated network. This allows us to use the methods proposed in Mai et al. (2015) to quickly estimate the RNMEV model on a real network.

We propose 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. We show that the RCNL allows to exhibit a more general correlation structure at each choice stage. We report estimation results and a prediction study for a network comprising more than 3000 nodes and 7000 links. The results show that the RCNL model yields sensible parameter estimates and the in-sample and out-of-sample fit are significantly better than the NRL model.

Keywords: Route choice model, recursive network MEV, integrated network, recursive cross-nested, value iterations, maximum likelihood estimation, cross-validation.

*Department of Computer Science and Operational Research, Université de Montréal and CIRRELT, Canada

1 Introduction

Discrete choice models are generally used for analyzing route choices in real networks. There are two main issues associated with the estimation of the parameters of such models. First, choice sets of paths are unknown to the analyst. Second, path utilities may be correlated, for instance, due to physical overlap in the network. In this paper we propose a general and operational representation of the recursive route choice models. Similar to Mai et al. (2015), the choice of path is modeled as a sequence of state choices and can be consistently estimated without sampling choice sets of paths, it is straightforward to apply for prediction and allows the choice model at each stage to be any member of the network MEV model (Daly and Bierlaire, 2006).

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 of successor states (instantaneous utility) and the expected maximum utility from the states to the destination (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 an infinite number of path alternatives. The RL model hence inherits the IIA property which is undesirable in a route choice setting (Mai et al., 2014). Recently, Mai et al. (2015) proposed the nested RL (NRL) model that relaxes the IIA property over paths by assuming that scale parameters are link specific. The model however assumes that the random terms of instantaneous utilities at each choice stage are i.i.d. extreme value type I and the choice at each stage is hence given by a logit model. So the correlations between the utilities of successor states cannot be captured.

In this paper, we generalize these works by allowing the choice model at each stage to be any member of the network MEV model. A computational advantage of the NRL model is that the value functions can be computed by using a value iteration method, which can be expressed as a sequence of matrix operations that are fast to compute. The network MEV model forms complicated expected maximum utility and choice probabilities. Therefore, in the case of the RNMEV, the value functions are a solution to a complex non-linear system which is substantially more difficult to deal with and the methods proposed in Mai et al. (2015) cannot be used directly. We deal with this challenge by using the NRL model but based on a new artificial network to simplify the estimation of the RNMEV. More precisely, the new network is created by integrating the networks of correlation structures given by the network MEV models into the road network and we show that the estimation of the RNMEV model can be simplified by applying the NRL model on the integrated network. The methods in Mai et al. (2015) then can be applied to obtain the value functions as well as the likelihood of the RNMEV model. Moreover, we also propose a recursive cross-nested logit (RCNL) model, a member of the RNMEV model, which allows the choice model at each stage to be a cross-nested logit model. We show that with the RCNL model, the variance-covariance matrix at each choice stage is no longer diagonal as in the RL or NRL model. This model therefore allows to exhibit a more general correlation structure at each choice stage, compared to the NRL and RL models.

This paper makes a number of contributions. First, we propose a general representation of the recursive route choice models that can be consistently estimated and used for prediction without sampling choice sets while allowing the random terms to be correlated. Second, we show that the generalized model can be estimated quickly by using the NRL model based on an integrated network. Third, we propose the RCNL model that can flexibly capture correlation structures at each choice stage. Fourth, we provide estimation and cross-validation results for a real network using simulated and real observations. Finally, the estimation code is implemented in MATLAB and is freely available upon request.

The paper is structured as follows. Section 2 presents the RNMEV model. Section 3 show the similarities between the RNMEV and RNL model. Section 4 discusses the maximum likelihood estimation. Section 5 presents the RCNL model. We present the numerical results in Section 6 and finally Section 7 concludes. The detailed proofs for the theorems proposed in the paper are provided in the Appendix.

2 Recursive network MEV model

In the RL model (Fosgerau et al., 2013a) and NRL model (Mai et al., 2015), 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 independently and identically distributed (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 allowing the choice at each stage to be a network MEV model. The model is derived based on a dynamic discrete choice model with a discount factor of 1 as in Mai et al. (2015). However, in order to better describe the 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. (2015) 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 Figure 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.

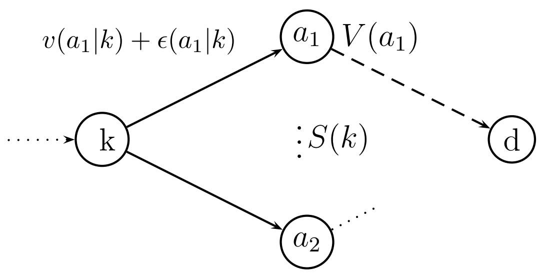

Figure 1: Illustration of notation

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 and random terms $(\epsilon(a|k), a \in S(k))$ follow a MEV distribution, with the generating function G_k of homogeneous degree $\mu_k > 0$ generated by the network MEV model (Daly and Bierlaire, 2006). Note that the term $\frac{\gamma}{\mu_k}$ is used in order to ensure that the random term has zero mean as in Fosgerau et al. (2013a), Mai et al. (2015). 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 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

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

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 as

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

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 (1978) the value functions can be computed as

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

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

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

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 $a, k \in \tilde{\mathcal{S}}$. Moreover, 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 Equation (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 generating function $G_k(y)$ is generated by a network MEV model based on a cycle-free graph 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. Here we remark that in order to be consistent we define the network MEV model based on states instead of nodes as in Daly and Bierlaire (2006). 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. Figure 2 shows a network of correlation structure at state k. Each arc $(i, j) \in \mathcal{A}_k$ is associated with a positive parameter α^k_{ij} and each state $i \in \mathcal{S}_k$ is associated with a positive scales ξ^k_i. The choice probability generating functions (CPGF) (Fosgerau et al., 2013b) (with respect to a vector of parameters y) associated with each state in $\mathcal{S}_k$ are defined as

$$G^i_k(y) = y_i^{\xi^k_i}, \quad i \in C_k, \tag{6}$$

and

$$G^i_k(y) = \sum_{j \in S_k(i)} \alpha^k_{ij} (G^j_k(y))^{\xi^k_i/\xi^k_j}, \forall i \in \mathcal{S}_k \backslash C_k, \tag{7}$$

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

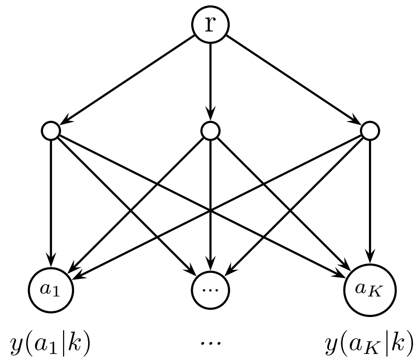

Figure 2: A network of correlation structure at state k

the CPGF $G_k(y)$ as $G_k(y) = G^r_k(y)$, where r is the root of network $\mathcal{G}_k$. Daly and Bierlaire (2006) show that $G_k(y) = G^r_k(y)$ is a ξ^k_r-MEV CPGF and μ_k is the homogeneous degree of $G_k(y)$, so $\xi^k_r = \mu_k$.

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

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

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 $a \in S(k)$.

Daly and Bierlaire (2006) show that 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 link-nested logit model (Vovsha and Bekhor, 1998) and the GenL model (Swait, 2001) models. So the RNMEV model allows to capture the correlations at each choice stage in many different ways. Indeed, if all the G_k functions, $\forall k \in \mathcal{S}$, refer to the MNL model, we obtain the NRL model, and the model can be estimated quickly for a large network using the methods presented in Mai et al. (2015).

Basically, function G_k and the choice probability $P_k(i|C_k)$ are complicated, the value functions become expensive to solve and we cannot use the methods in Fosgerau et al. (2013a) or Mai et al. (2015). The gradients of the value functions are also cumbersome to evaluate as well. In the next section we show how to simplify the estimation of the model by integrating the networks of correlation structures $\mathcal{G}_k$ $\forall k \in \mathcal{S}$ into the road network $\mathcal{G}$. This allows to use the estimation methods proposed in Mai et al. (2015) to estimate the RNMEV model.

3 Integrated network and similarities between the NRL and RNMEV models

In this section we show how to integrate the network $\mathcal{G}_k$, $\forall k \in \mathcal{S}$, into the network $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ in order to simplify the estimation of the model. More precisely, we first introduce a method to integrate the networks of correlation structures into the road network, second we show that the estimation of the RNMEV model can be done by using the value functions and choice probabilities given by the NRL model based on the integrated network.

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 in the above section, 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 state $k \in \mathcal{S}$, 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 $S^*(k)$ be the set of successor states of state k in network $\mathcal{G}^*$.

Note that due to the properties of the network $\mathcal{G}_k$ (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 Figure 3 a small example where state k has three successors a_1, a_2 and a_3 as illustrated in the left part of Figure 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.

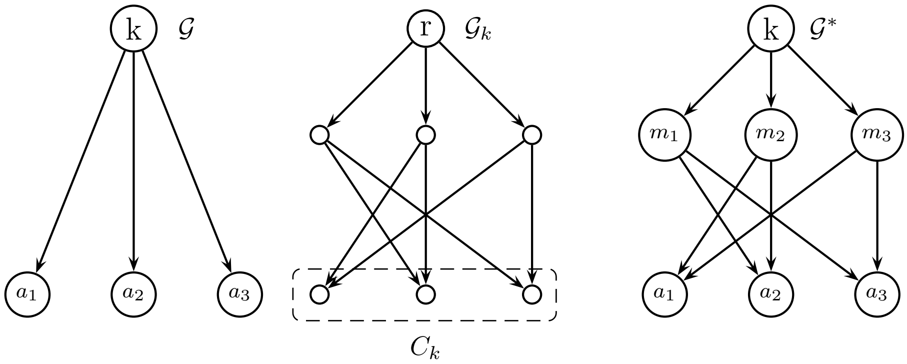

Figure 3: Original and integrated network

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 exits 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 (vi) 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. ■

3.2 Integrated network with NRL model

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.

Now we apply the NRL model (Mai et al., 2015) 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). Note 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. (2015).

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. (2015), 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 the NRL on network $\mathcal{G}^*$ and 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*

$$Y^*_k = G_k\Big(e^{v(a|k)}(Y^*_a)^{1/\mu_a}, a \in S(k)\Big), \ \forall k \in \mathcal{S}. \tag{17}$$

In other words, $Y_k = Y^*_k$, $\forall k \in \tilde{\mathcal{S}}$, is a solution to the Bellman's equation of the RNMEV model in (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*

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

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

We provide the proofs of the two theorems in Appendixes 1 and 2. Theorems 1 and 2 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 $\mathcal{G}^*$. So the methods proposed in Mai et al. (2015) can be used. In the next section we discuss in detail the maximum likelihood estimation.

4 Maximum likelihood estimation

Aguirregabiria and Mira (2010) discuss different ways of estimating a dynamic discrete choice model. Similar to Fosgerau et al. (2013a) and Mai et al. (2015) the nested fixed point algorithm (Rust, 1987) can be used to estimate the RNMEV 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. (2015), namely estimating the value functions with dynamic accuracy and computing the gradient 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. (2015), we define a matrix $M^*(|\mathcal{S}^*| \times |\mathcal{S}^*|)$ with entries

$$M^*_{ka} = \delta^*(a|k) e^{\mu^*_k v^*(a|k)} \, \forall k, a \in \mathcal{S}^*, \tag{19}$$

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 $||(Y^*)^{i+1} - (Y^*)^i|| \leq \gamma$, for a given threshold $\gamma > 0$ as stopping criteria. Mai et al. (2015) 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. (2015) 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 under the RNMEV model. The LL function defined over the set of path observations $n = 1, \dots, N$ 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 $\mathcal{G}$. And for 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 result of Theorems 1 and 2. This leads to the fact that the LL function becomes more complicated than the one given in Mai et al. (2015)

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

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

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 gradient of the LL function is complicated, but can be easily derived based on (22), (2) and (16). Its requires 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. (2015) 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 homoschedastic variance-covariance matrix. 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 and Bierlaire (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

$$G_k(y) = \sum_{m \in S_k(r)} \Big(\sum_{a \in S_k(m)} \alpha^k_{ma} y_a^{\xi^k_m} \Big)^{\xi^k_r/\xi^k_m}, \tag{25}$$

(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 correlations between the successor states $a \in S(k)$, we define 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. Figure 4 shows 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}, $\forall i, j = 1, 2, 3$.

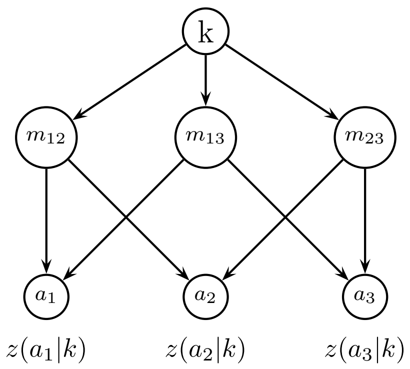

Figure 4: A cross-nest logit model structure at link k

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

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

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

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

So that 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 $Corr\left(z(a_i|k), z(a_j|k)\right) = 0$, 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 correlations 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 NRL model to 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

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

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

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

and

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

6 Numerical results

For the numerical results we use the same data as Fosgerau et al. (2013a), Mai et al. (2015) (also used in Frejinger and Bierlaire, 2007) 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., 2015).

6.1 Model specification

We use the same attributes as Mai et al. (2015) for the instantaneous utilities. We note that Mai et al. (2015) 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. First, travel time $TT(a)$ of action a. Second, a left turn dummy $LT(a|k)$ that equals one if the turn angle from k to a is larger than 40 degrees and less than 177 degrees. Third, a u-turn dummy $UT(a|k)$ that equals one if the turn angle is larger than 177. Fourth, a state constant $LC(a)$. 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, $k \in \mathcal{S}, a \in S(k)$

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

In the NRL model in Mai et al. (2015), the scale of the random terms are defined as exponential functions of the model parameters. Mai et al. (2015) 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. This fairly small 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. (2015) 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, Equation 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

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

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

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

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 be positive i.e. $\lambda_{TT}, \lambda_{LS}, \lambda_{OL} \geq 0$.

In the CNL model at state k, parameter α_{ma}^k reflects the level of membership of alternative a to nest m. 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 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

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

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

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

where $\epsilon(a)$ is i.i.d extreme value type I and $\epsilon(a|k;\omega)$, $a \in S(k)$, have a 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. Interestingly, 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 correlations between successor states.

The final log-likelihood 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. (2015), we observe a significant improvement in final log-likelihood 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. Interestingly, the final-log likelihood 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/links and the integrated network has 31373 states, so there are 24084 new states added to the original network. The number of arcs in the integrated network is 72252, compared to 20202 arcs in the original network. So the integrated network is larger than the original, leading to the fact that 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. We note that 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. (2015), we use a cross validation approach i.e. the sample of observations is divided into two sets by drawing observations uniformly with a fixed probability: one set (80% of the observations) is used for estimation and the other (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 log-likelihood loss to evaluate the prediction performance.

For each holdout sample i, $0 \leq i \leq 20$, we estimate the parameters $\hat{\beta}_i$ off the corresponding training sample 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 i. Then err_i is a random variable that depends on the holdout sample i. In order to have unconditional test error values we compute the average of err_i values over samples as follows

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

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 1: Estimation 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

Table 2: Likelihood ratio test results

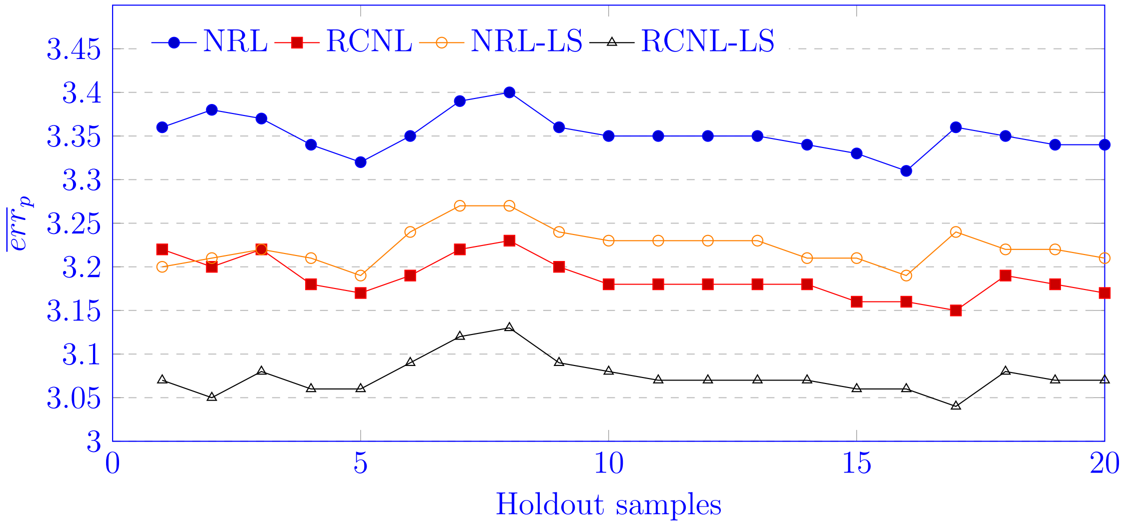

Figure 5: Average of the test error values over holdout samples

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

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

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

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.

7 Conclusion

This paper has presented a general and operational representation of the recursive route choice models. The RNMEV model is an extension of the NRL model proposed in Mai et al. (2015) where the choice of each stage is a network MEV model. We have showed that the model can be estimated by applying the NRL model to a new network which is created by iterating the networks of correlation structures at each choice stage into the road network. So the methods proposed in Mai et al. (2015) can be used to estimate the new model.

We have proposed the RCNL model, a member of the RNMEV model, by allowing the model at each choice stage to be a CNL model. We showed that the RCNL can exhibit a more general correlation structure at each choice stage, compared to the NRL and RL models.

We have provided numerical results using a real data. 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-LS and NRL models.

In this paper we use a 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 network where the correlations at each choice stage need to be taken into account seriously e.g. dynamic networks (state is time and location) and multi-modal networks (state is location and mode).

Finally, we note that the methods proposed in this paper are not restricted to route choice applications. By adapting the network of state space we can deal with the estimation of complicated dynamic discrete choice models where the choice at each stage can be the network MEV model instead of MNL.

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 discussions with Fabian Bastin and Emma Frejinger.

A A proof for 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 $G_k\Big(e^{v(a|k)}(Y_a^*)^{1/\mu_a}, a \in S(k)\Big)$ 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^* be a vector of size $|S(k)|$ with entries $e^{v(a|k)}(Y_a^*)^{1/\mu_a}$, for all $a \in S(k)$. 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 $\mathcal{L}^k(i)$ be 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 $\mathcal{T}^k(p)$ 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 exits a number $p' \geq p$ such that $j \in \mathcal{T}^k(p')$. It means that there exits 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 exits 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 in the following

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

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

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

Substitute (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, Equation 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 $p = I \geq 1$. In other words

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

For each state $i \in \mathcal{T}^k(I+1)$, according to Proposition 2(iv) if $j \in S^*(i)\backslash S(k)$ then $j \in \bigcup_{t=1}^{I} \mathcal{T}^k(t)$. Consequently, by assumption, $G_k^j(y_k^*) = Y_j^*$ $\forall j \in S^*(i)\backslash S(k)$. Equation 39 can be written as

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

so $G_k^i(y_k^*) = Y_i^*$, $\forall i \in \mathcal{T}^k(I+1)$, because of (15). This validates (36) for $p = I+1$. So the lemma is completely proved. $\blacksquare$

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

B A proof for 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

$$P(a|k) = \sum_{[j_0,\dots,j_I]\in\Omega^k(a)} \prod_{t=0}^{I-1} \frac{\alpha^k_{j_t j_{t+1}} (G_k^{j_{t+1}}(y_k^*))^{\xi^k_{j_t}/\xi^k_{j_{t+1}}}}{G_k^{j_t}(y_k^*)} \tag{42}$$

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 Equation 16 we obtain
$$\frac{\alpha^k_{ij}(G_k^j(y_k^*))^{\xi^k_i/\xi^k_j}}{G_k^i(y_k^*)} = e^{\mu_i^* v^*(j|i)} \frac{(Y_j^*)^{\mu_i^*/\mu_j^*}}{Y_i^*} = P^*(j|i). \tag{43}$$
- If $j \in S(k)$, from (37) we have
$$\begin{aligned}\frac{\alpha^k_{ij}(G_k^j(y_k^*))^{\xi^k_i/\xi^k_j}}{G_k^i(y_k^*)} &= \frac{e^{\mu_i^*(v^*(j|i)-v(j|k))}}{Y_i^*} e^{\mu_i^* v(j|k)}(Y_j^*)^{\mu_i^*/\mu_j^*} \\ &= e^{\mu_i^* v^*(j|i)} \frac{(Y_j^*)^{\mu_i^*/\mu_j^*}}{Y_i^*} = P^*(j|i).\end{aligned} \tag{44}$$

So from (43) and (44) we get

$$\frac{\alpha^k_{ij}(G_k^j(y_k^*))^{\xi^k_i/\xi^k_j}}{G_k^i(y_k^*)} = P^*(j|i),\ \forall i \in \mathcal{S}_k \backslash S(k), j \in S^*(i)$$

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

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

Hence, the theorem is proved.

References

Abbé, E., Bierlaire, M., and Toledo, T. Normalization and correlation of cross-nested logit models. *Transportation Research Part B*, 41(7):795–808, 2007.

Aguirregabiria, V. and Mira, P. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67, 2010.

Ben-Akiva, M. and Bierlaire, M. Discrete choice methods and their applications to short-term travel decisions. In Hall, R., editor, *Handbook of Transportation Science*, pages 5–34. Kluwer, 1999.

Ben-Akiva, M. *The structure of travel demand models.* PhD thesis, MIT, 1973.

Daly, A. and Bierlaire, M. A general and operational representation of generalised extreme value models. *Transportation Research Part B: Methodological*, 40(4): 285 – 305, 2006.

Fosgerau, M., Frejinger, E., and Karlström, A. A link based network route choice model with unrestricted choice set. *Transportation Research Part B*, 56:70–80, 2013a.

Fosgerau, M., McFadden, M., and Bierlaire, M. Choice probability generating functions. *Journal of Choice Modelling*, 8:1–18, 2013b.

Frejinger, E. and Bierlaire, M. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B*, 41(3):363–378, 2007.

Koppelman, F. and Wen, C.-H. The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B*, 34:75–89, 2000.

Mai, T., Frejinger, E., and Bastin, F. A misspecification test for logit based route choice models. *Technical report, CIRRELT-2014-32*, 2014.

Mai, T., Fosgerau, M., and Frejinger, E. A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, 75(0):100 – 112, 2015. ISSN 0191-2615.

McFadden, D. Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Residential Location*, pages 75–96. North-Holland, Amsterdam, 1978.

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

Rust, J. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033, 1987.

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

Swait, J. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B*, 35:643–666, 2001.

Vovsha, P. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record*, 1607:6–15, 1997.

Vovsha, P. and Bekhor, S. Link-nested logit model of route choice Overcoming route overlapping problem. *Transportation Research Record*, 1645:133–142, 1998.

Wen, C.-H. and Koppelman, F. The generalized nested logit model. *Transportation Research Part B*, 35(7):627–641, 2001.