6-2017

On the similarities between random regret minimization and mother logit: The case of recursive route choice models

Tien MAI
Singapore Management University, atmai@smu.edu.sg

Fabian BASTIN

Emma FREJINGER

Journal of Choice Modelling 23 (2017) 21–33

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

On the similarities between random regret minimization and mother logit: The case of recursive route choice models

Tien Mai[a,*], Fabian Bastin[b], Emma Frejinger[b]

[a] Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal and CIRRELT, Canada
[b] Department of Computer Science and Operational Research, Université de Montréal and CIRRELT, Canada

ARTICLE INFO

Keywords:
Route choice modeling
Recursive logit
Random regret minimization
Mother logit
Maximum likelihood estimation
Cross-validation

ABSTRACT

This paper focuses on the comparison of the random regret minimization (RRM) and mother logit models for analyzing the choice between alternatives having deterministic attributes. The mother logit model allows utilities of a given alternative to depend on attributes of other alternatives. It was designed to relax the independence from irrelevant alternatives (IIA) property while keeping the random terms independently and identically distributed extreme value distributed (McFadden et al., 1978).

We adapt and extend the RRM model proposed by Chorus (2014) to the case of recursive logit (RL) route choice models (Fosgerau et al., 2013). We argue that these RRM models can be cast as mother logit models and we define such models that are equivalent to the RRM ones considered in this paper. The results show that one of the RRM models and its mother logit equivalent has the best out-of-sample fit indicating that utility functions based on attribute differences best explains the choices in our application.

1. Introduction

Regret theory was introduced decades ago (e.g. Loomes and Sugden, 1982) and similar to prospect theory (Kahneman and Tversky, 1979), it was originally designed for modeling choice under uncertainty. Chorus (2010) linked the theory to discrete choice modeling and proposed a random regret minimization (RRM) model that can be estimated by maximum likelihood using standard software. He derived the model based on the assumption that decision-makers try to avoid the situation where a non-chosen alternative outperforms a chosen one in terms of observed attributes. Several extensions to this model have been proposed (e.g. Chorus, 2012, 2014). Unlike the original theory, the applications of these recent models have been limited to deterministic alternatives, i.e. to known attribute values.

These RRM models relax the independence of irrelevant alternatives (IIA) property even though the random terms are independently and identically distributed (i.i.d.) extreme value type I. This is due to the fact that the regret associated with an alternative depends on the attributes of other alternatives. In this context, the regret models share similarities with mother logit ones, which have been designed to relax the IIA property from logit models (McFadden, 2000; McFadden et al., 1978).

In this paper we focus on the comparison between the RRM and mother logit for modeling the choice of alternatives having deterministic attributes. We do this in the context of route choice modeling and our RRM models are based on the Generalized Random Regret Minimization (GRRM) model proposed by Chorus (2014).

Discrete choice models are widely used for analyzing path choices in real networks. Following the discussion in Fosgerau et al.

* Corresponding author.
E-mail address: AnhTien.Mai@cirrelt.ca (T. Mai).

http://dx.doi.org/10.1016/j.jocm.2017.03.002
Received 14 July 2015; Received in revised form 10 December 2016; Accepted 6 March 2017
Available online 04 April 2017
1755-5345/

(2013), route choice models in the literature can be grouped into three approaches. First, the classical path-based approach where choice sets of paths are generated and treated as the actual choice sets. Second, the sampling approach (Frejinger et al., 2009; Lai and Bierlaire, 2015) where choice sets of paths are sampled and utilities are corrected for the sampling protocol so that the estimator is consistent. Fosgerau et al. (2013) proposed a third link-based approach that allows to consistently estimate parameters and forecast path choices without sampling any choice sets of paths. The link-based model is called recursive logit and is used in this paper. Mai et al. (2015a) proposed a nested recursive logit model where random terms are correlated. Both models are based on the random utility maximization (RUM) framework.

Prato (2014) analyzes path-based route choice model estimation results using the RRM model proposed by Chorus (2010). He focuses on the two well-known challenges associated with route choice modeling, namely, choice set generation and correlation. He observes that RRM models perform well on real data, but in an experimental setting, he finds that the parameter estimates of RRM models have the wrong signs when so-called "irrelevant alternatives" are included in the choice sets. Following the finding in Horowitz and Louviere (1995), the RL model is based on the universal choice set of all paths connecting an origin-destination pair and does not need choice set generation. We investigate whether the RL model presents similar issues as the one considered in Prato (2014) and we analyze the out-of-sample fit.

This paper makes a number of contributions linking RRM to mother logit models. We propose two specifications for RL models with random regret. The first model (called Extended Random Regret Minimization - ERRM) extends the GRRM model by adding terms associated with the attributes of the non-chosen alternatives to the attribute differences in the regrets. The second model, called Averaged Random Regret Minimization (ARRM) model, modifies the first one by averaging the regrets over the alternatives. We prove that under some constraints on the parameters, the choice probabilities given by the ARRM are equivalent to those given by the RUM model proposed in Fosgerau et al. (2013). This model therefore generalizes the RUM-based RL model. Furthermore, we propose three mother logit models that are equivalent to the RRM models considered in this paper. Finally, we report estimation and cross-validation results for a real network with over 3000 nodes and 7000 links. The estimation code for the RRM and mother logit RL models is implemented in MATLAB and we share the code freely on GitHub as an open source project.[1]

The paper is structured as follows. In Section 2 we review the RUM, mother logit and RRM models. In Section 3 we propose the RL model under RRM, and Section 4 presents two different formulations for link regrets. The RUM models that are equivalent to the RRM ones are presented in Section 5. In Section 6 we discuss maximum likelihood estimation. Model specifications as well as estimation and cross-validation results are presented in Section 7, and finally, Section 8 concludes.

2. Random utility maximization, mother logit and random regret minimization models

In this section we start by introducing the modeling ideas behind the RL model followed by a brief review of RUM, mother logit and RRM models. The RRM RL model is presented in more detail in Section 3.

The RL model is based on the dynamic discrete choice framework proposed by Rust (1987). The path choice problem is formulated as a sequence of link choices, and at each stage the travelers choose a next link by observing the costs of the outgoing links and the expected costs from the current state to the destination. The link choice probabilities can then be computed by defining link costs and a decision rule at each choice stage. In Fosgerau et al. (2013), the RUM decision rule is used at each stage, i.e., they assume that the traveler aims at maximizing the sum of the random utility of outgoing links (instantaneous utility) and the expected maximum utility from the sink node of the links to the destination (value function). The random terms of the instantaneous utilities are assumed to be i.i.d. extreme value type I and the RL model is equivalent to a logit model over the sets of all the feasible path alternatives.

In the RUM discrete choice models, an individual n associates a utility U_{ni} with an alternative i within a choice set C_n. The utility consists of two terms $U_{ni} = V_{ni} + \epsilon_{ni}$: a deterministic V_{ni} part, observed by the modeler, and a random part ϵ_{ni}. Typically, V_{ni} is a linear-in-parameters function of attributes, i.e., $V_{ni} = \beta^T x_{ni}$, where β is a vector of parameters to be estimated and x_{ni} is a vector of attributes with respect to individual n and alternative i. A decision-maker chooses the alternative that maximizes his/her utility

$$i^* = \underset{i \in C_n}{\operatorname{argmax}} \{V_{ni} + \epsilon_{ni}\}.$$

The multinational logit (MNL) model is based on the assumption that the random terms ϵ_{ni} are i.i.d. extreme value type I, and the probability of choosing an alternative i is $P_n(i) = e^{V_{ni}} / \sum_{j \in C_n} e^{V_{nj}}$.

The MNL model, however, retains the IIA property, so other models may be preferred in order to capture the correlation between random terms e.g. the nested logit model (Ben-Akiva, 1973), cross-nested logit model (Vovsha and Bekhor, 1645) or network multivariate extreme value model (Daly and Bierlaire, 2006).

The mother logit model (also called universal logit model) was introduced decades ago by McFadden et al. (1978). It was designed to relax the IIA property of the logit model while keeping the random terms i.i.d. extreme value, and can approximate any discrete choice model that is continuous in its arguments on a compact set (McFadden, 1984). The mother logit approximation however does not require that the discrete choice model comes from a RUM model, while McFadden and Train (2000) show that under mild regularity conditions, mixed logit models are RUM models, and any RUM discrete choice model can be approximated as closely as desired by a mixed logit model. In this paper we show that we can define mother logit models that are equivalent to RRM models.

[1] https://github.com/maitien86/

It is interesting to note that it is common practice in route choice analysis to use logit models that include attributes in the utility of a given alternative that depend on characteristics of other alternatives. Examples are path size logit (Ben-Akiva and Bierlaire, 1999), c-logit (Cascetta et al., 1996) and link size (Fosgerau et al., 2013). In this case, the resulting model is actually mother logit and it may not be consistent with RUM.

The RRM models are based on the assumption that decision-makers try to avoid the situation where a non-chosen alternative outperforms a chosen one in terms of one or more attributes. This translates into a regret function for a considered alternative by including all attributes of all competing alternatives. The random regret RR_{ni} can be written as the sum of a deterministic part R_{ni} and a random error term ϵ_{ni} (Chorus, 2012),

$$RR_{ni} = R_{ni} + \epsilon_{ni} = \sum_{j\neq i, j\in C_n} \sum_t \ln\left(1 + e^{\beta_t(x_{nj}(t) - x_{ni}(t))}\right) + \epsilon_{ni}, \tag{1}$$

where t is an attribute. The regret R_{ni} is thus computed based on two sums, the first is over all other alternatives in the choice set and the second over all the attributes. Contrary to the RUM models, a decision-maker aims to minimize the random regret

$$i^* = \underset{i\in C_n}{\operatorname{argmin}}\{R_{ni} + \epsilon_{ni}\} = \underset{i\in C_n}{\operatorname{argmax}}\{-R_{ni} - \epsilon_{ni}\}. \tag{2}$$

Under the assumption that the random terms $-\epsilon_{ni}$ are i.i.d extreme value type I, the choice probability is given by the MNL model

$$P_n(i) = \frac{e^{-R_{ni}}}{\sum_j e^{-R_{nj}}}.$$

It is important to note that even though this is the logit model, the IIA property does not hold since the regrets are not alternative specific. Chorus (2014) presents the Generalized Random Regret Minimization (GRRM) model, where the random regret can be expressed as

$$GRR_{ni} = \sum_{j\neq i, j\in C_n} \sum_t \ln\left(\lambda_t + e^{\beta_t(x_{nj}(t) - x_{ni}(t))}\right) + \epsilon_{ni}. \tag{3}$$

We note that $GRR_{ni} = RR_{ni}$ if $\lambda_t = 1,\ \forall t$. Moreover, as pointed out in Chorus (2014), if $\lambda_t = 0\ \ \forall t$ the resulting regret becomes linear-in-parameters

$$GRR_{ni} = \sum_{j\neq i, j\in C_n} \sum_t \beta_t(x_{nj}(t) - x_{ni}(t)) = \sum_{j\in C_n} \beta^T x_{nj} - |C_n|\beta^T x_{ni},$$

where $|\cdot|$ is the cardinality operator. The term $\sum_{j\in C_n} \beta^T x_{jn}$ is the same for any alternative i, and does not affect the choice given by (2). The regret has a linear-in-parameters formulation but it is different from the RUM model because of $|C_n|$.

A disadvantage of the RRM or GRRM model, highlighted in Chorus (2012), is that the running time for computing the choice probabilities increases exponentially with the size of the choice set. Indeed, every alternative is compared with every other in terms of each attribute. This can hence be a problem for path-based route choice applications which are characterized by large choice sets. This is not an issue for the link-based RL model that we present in the following section since the choice set at each stage is small: equal to the number of outgoing links at a node.

3. Random regret recursive logit models

In the RUM recursive logit model (RL-RUM) proposed by Fosgerau et al. (2013), a linear-in-parameter utility is associated with each link pair in the network, and is the sum of a deterministic and a random term. A traveler maximizes his/her utility, defined as the sum of the instantaneous link utility at the current decision stage and the expected maximum utility from the sink node of outgoing links to the destination. The random terms are assumed to be i.i.d. extreme value type I, so the choice model at each stage is MNL, leading to the fact that the expected maximum utilities can be computed by solving a system of linear equations. In the following we present a RL model based on the RRM decision rule. The derivation is similar to Fosgerau et al. (2013) but the link costs and expected costs are different.

A directed connected graph (not assumed acyclic) $\mathcal{G} = (\mathcal{A}; \mathcal{V})$ is considered, where $\mathcal{A}$ and $\mathcal{V}$ are the set of links and nodes, respectively. For each link $k \in \mathcal{A}$, we denote the set of outgoing links from the sink node of k by $A(k)$. We extend the network with a dummy link d, without successors, per destination, that is, an absorbing state. The set of all links for a given destination is hence $\widetilde{\mathcal{A}} = \mathcal{A} \cup \{d\}$. Given two links $a, k \in \widetilde{\mathcal{A}}$, $a \in A(k)$, we associate the following instantaneous random regret for individual n

$$rr_n(a|k) = r_n(a|k) + \mu\epsilon_n(a),$$

where $r_n(a|k)$ is the deterministic part of the link regret associated with link a given k, $-\epsilon_n(a)$ are i.i.d. extreme value type I distributed error terms and μ is a strictly positive scale parameter. We ensure that $\epsilon_n(a)$ have zero mean by adding Euler's constant. For notational simplicity, we omit from now on the index for individual n but note that the regrets and random terms can be individual specific.

At each state k the traveler observes the realizations of the random terms $\epsilon(a)$, $a \in A(k)$. He/she then chooses a link $a \in A(k)$ that minimizes the sum of instantaneous random regret $r(a|k)$ and expected downstream regret. The latter, denoted by $R^d(k)$, is defined as

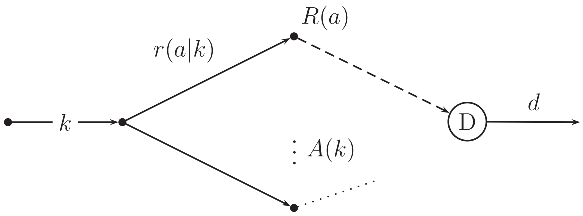

Fig. 1. Illustration of notation.

the expected minimum regret from state k to the destination (see Fig. 1). The superscript d indicates that the expected minimum regrets are destination specific (through dummy link d). $R^d(k)$ is recursively defined by the Bellman equation as

$$R^d(k) = \mathbb{E}\left[\min_{a \in A(k)} \{r(a|k) + R^d(a) + \mu\epsilon(a)\}\right], \quad \forall\, k \in \mathcal{A}. \tag{4}$$

We note that $R^d(k)$ and $r(a|k)$ may be conditional on the model parameters so they can be written as $R^d(k) = R^d(k;\beta)$ and $r(a|k) = r(a|k;\beta)$ where β is the vector of parameters to estimate. We, however, omit β for notational simplicity. Eq. (4) can be written as

$$R^d(k) = \mathbb{E}\left[-\max_{a \in A(k)} \{-r(a|k) - R^d(a) - \mu\epsilon(a)\}\right] = -\mathbb{E}\left[\max_{a \in A(k)} \{-r(a|k) - R^d(a) + \mu(-\epsilon(a))\}\right], \quad \forall\, k \in \mathcal{A},$$

or equivalently

$$\frac{1}{\mu}R^d(k) = -\mathbb{E}\left[\max_{a \in A(k)}\left\{\frac{1}{\mu}(-r(a|k) - R^d(a)) + (-\epsilon(a))\right\}\right], \quad \forall\, k \in \mathcal{A} \tag{5}$$

Since $-\epsilon(a)$ are i.i.d. standard extreme value type I by assumption, the probability of choosing link a given k is given by the MNL model

$$P^d(a|k) = \frac{\xi(a|k)e^{-\frac{1}{\mu}(r(a|k)+R^d(a))}}{\sum_{a' \in A(k)} e^{-\frac{1}{\mu}(r(a'|k)+R^d(a'))}}, \quad \forall\, a, k \in \widetilde{\mathcal{A}}. \tag{6}$$

Note that we include $\xi(a|k)$ that equals one if $a \in A(k)$ and zero otherwise so that the probability is defined for all $a, k \in \widetilde{\mathcal{A}}$ (we recall that $\widetilde{\mathcal{A}} = \mathcal{A} \cup \{d\}$). Since the choice model at each state is MNL, the expected minimum regrets are given recursively by the logsum

$$-\frac{1}{\mu}R^d(k) = \mathbb{E}\left[\max_{a \in A(k)}\left\{\frac{1}{\mu}(-r(a|k) - R^d(a)) + (-\epsilon(a))\right\}\right] = \ln\left(\sum_{a \in A(k)} e^{\frac{1}{\mu}(-r(a|k) - R^d(a))}\right), \quad \forall\, k \in \mathcal{A}, \tag{7}$$

and $R^d(d) = 0$ by assumption. We define a matrix M^d of size $|\widetilde{\mathcal{A}}| \times |\widetilde{\mathcal{A}}|$ and a vector z of size $|\widetilde{\mathcal{A}}|$ with entries

$$M^d_{ka} = \xi(a|k)e^{-\frac{1}{\mu}r(a|k)}, \quad z^d_k = e^{-\frac{1}{\mu}R^d(k)}, \quad \forall\, k, a \in \widetilde{\mathcal{A}}. \tag{8}$$

From (7) we have

$$z^d_k = \begin{cases} \sum_{a \in \mathcal{A}} M^d_{ka} z^d_a & \text{if } k \in \mathcal{A} \\ 1 & \text{if } k = d, \end{cases} \tag{9}$$

and the system in (9) can be written in matrix form as

$$z^d = M^d z^d + b,$$

or equivalently,

$$z^d = (I - M^d)^{-1} b, \tag{10}$$

where b is a vector of size $|\widetilde{\mathcal{A}}|$ with zeros values for all states except for the destination d that equals 1 and I is the identity matrix. Similar to Fosgerau et al. (2013) we obtain a system of linear equations which can be solved in short computational time. Fosgerau et al. (2013) discuss the existence of a solution to the Bellman equation for the RL-RUM model and this can be applied in the context of the RRM-based RL models. In essence, the existence of a solution depends on the size of the scaled instantaneous regrets and on the balance between the number of paths connecting the nodes in the network. It is easy to find a feasible solution by increasing the magnitude of the parameters. Note that if the scales μ are different over links, the system in (9) becomes nonlinear, similar to the nested recursive logit model described in Mai et al. (2015a).

Using (6)and(7), the probability of choosing link a given a state k can be written as

$$P^d(a|k) = \xi(a|k)e^{-\frac{1}{\mu}(r(a|k)+R^d(a)-R^d(k))}, \quad \forall\, k, a \in \widetilde{\mathcal{A}},$$

and the probability of a path defined by a sequence of links $\sigma = [k_0, \dots, k_J]$ is

$$P^d(\sigma) = \prod_{i=0}^{J-1} P^d(k_{i+1}|k_i) = e^{\frac{1}{\mu}R^d(k_0)} \prod_{i=0}^{J-1} e^{-\frac{1}{\mu}r(k_{i+1}|k_i)} = e^{\frac{1}{\mu}R^d(k_0)} e^{-\frac{1}{\mu}r(\sigma)},$$

where $r(\sigma) = \sum_{i=0}^{J-1} r(k_{i+1}|k_i)$. Given two paths σ_1 and σ_2, the ratio between two probabilities is

$$\frac{P(\sigma_2)}{P(\sigma_2)} = e^{\frac{1}{\mu}(r(\sigma_2)-r(\sigma_1))},$$

and it does not only depend on the attributes of links on paths σ_1, σ_2. Hence, the IIA property does not hold for the RRM models. In the next section we discuss different formulations of the instantaneous link regret functions.

4. Link regret formulations

We define the regret $r(a|k)$ of link $a \in A(k)$ conditional on link $k \in \mathcal{A}$, based on the GRRM model given by (3). We start by noting that (3) is undefined when the choice set C_n is singleton. This can be an issue if a transport network has links with only one successor ($|A_k| = 1$). These links could be removed by pre-processing the network but we choose to derive link regret formulations based on all alternatives so that pre-processing is avoided. Indeed, if there is only one successor, the next-link probability is equal to 1 so that the path choice probability is unaffected. The slightly modified GRRM is

$$r^{\text{GRRM}}(a|k) = \sum_{a' \in A(k)} \sum_t \ln(\lambda_t + e^{\beta_t(x(a'|k)_t - x(a|k)_t)}), \quad \forall\, k \in \mathcal{A},\ a \in A(k), \tag{11}$$

where $x(a|k)$ is a vector of attributes associated with link a given k, λ and β are vector of parameters to be estimated. The only difference here with respect to the model in Chorus (2014) is that the first sum is over all alternatives, instead of all other alternatives.

We also define a new formulation for regret that we call Extended Random Regret Minimization (ERRM), by adding terms associated with the attributes of the non-chosen alternatives to the attribute differences in (11). The ERRM has the following formulation

$$r^{\text{ERRM}}(a|k) = \sum_{a' \in A(k)} \sum_t \ln(\lambda_t + e^{\beta_t(x(a'|k)_t - x(a|k)_t) + \delta_t x(a'|k)_t}), \quad \forall\, k \in \mathcal{A}, \quad a \in A(k) \tag{12}$$

and the difference with respect to GRRM lies in the term $\delta_t x(a'|k)_t$. If $\delta_t > 0$, the impact of the non-chosen alternatives becomes larger and if $\delta_t < 0$, it is smaller. Moreover, if $\delta_t = 0$ we obtain the GRRM formulation.

The regret in (12) can be written as

$$r^{\text{ERRM}}(a|k) = \sum_{a' \in A(k)} \sum_t \ln(\lambda_t + e^{-\beta_t x(a|k)_t + (\beta_t + \delta_t) x(a'|k)_t}), \quad \forall\, k \in \mathcal{A},\ a \in A(k),$$

which clearly shows that if $\lambda_t = 0$ and $\delta_t = -\beta_t, \forall\, t$, then the regret in (12) is linear-in-parameters and the attributes associated with the non-chosen alternatives cancel out. That is

$$r^{\text{ERRM}}(a|k;\ \lambda = 0,\ \delta = -\beta) = -|A(k)|\beta^T x(a|k) = -|A(k)| v(a|k),$$

where $v(a|k)$ are the linear-in-parameters utilities as in Fosgerau et al. (2013). In this case, the regret is linear-in-parameters but different from the RUM-based model with a factor $|A(k)|$. This factor appears because the sum in the regret formula is over all the outgoing links from the sink node of k. Therefore, we propose an Averaged Random Regret Minimization (ARRM) model as an alternative to the ERRM where a normalization factor is used so that the regret is averaged over all the alternatives

$$r^{\text{ARRM}}(a|k) = \frac{1}{|A(k)|} r^{\text{ERRM}}(a|k), \quad \forall\, k \in \mathcal{A},\ a \in A(k). \tag{13}$$

Moreover, we fix $\lambda_t = 0$ and $\delta_t = -\beta_t,\ \forall\, t$ to obtain $r^{\text{ARRM}}(a|k) = -v(a|k)$. Based on (8) the entries of matrix M^d becomes in this case

$$M^d_{ka} = \xi(a|k) e^{\frac{1}{\mu}v(a|k)}, \quad \forall\, k, a \in \widetilde{\mathcal{A}}.$$

We refer to the definition of the matrix M^d in Fosgerau et al. (2013) and note that z^d is a solution to the system of linear equations $(I - M^d)z^d = b$, therefore it is straightforward to show that

$$z^d_k = e^{-\frac{1}{\mu}R^d(k)} = e^{\frac{1}{\mu}V^d(k)}, \quad \forall\, k \in \widetilde{A},$$

where $V^d(k)$ is the expected maximum utility from state k to the destination. The probability of choosing a link a given link k can be written as

$$P^d(a|k) = \xi(a|k) e^{-\frac{1}{\mu}(r(a|k) + R^d(a) - R^d(k))} = \xi(a|k) e^{\frac{1}{\mu}(v(a|k) + V^d(a) - V^d(k))}.$$

This choice probability is equivalent to the one given by the RL-RUM model. So the RL model based on ARRM generalizes the RL-RUM model.

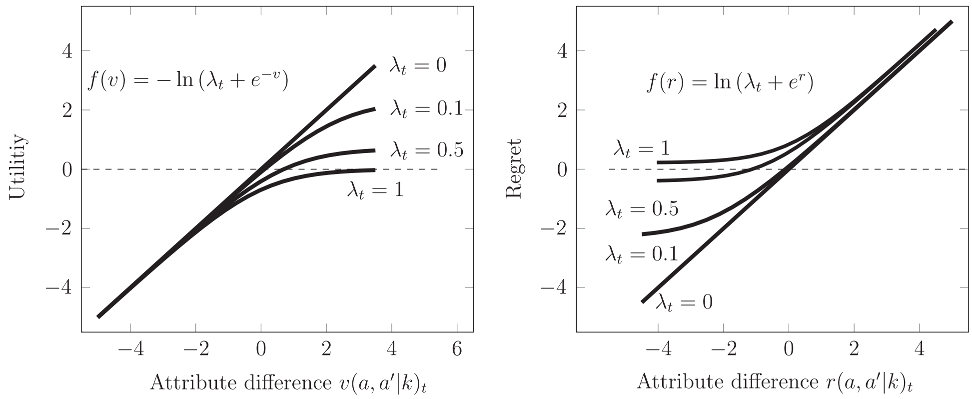

Fig. 2. Utilities and regrets given by different values of λ_t.

5. Equivalent mother logit models

Under the RUM framework, the utility associated with each alternative can only depend on the characteristics of that alternative. This is the case of the utility functions used in the different recursive logit models (Fosgerau et al., 2013; Mai, 2016; Mai et al., 2015a, 2016). The GRRM and ERRM link regret functions presented in the previous section include attributes from other alternatives in the choice set. These models are therefore similar to mother logit models with the only difference being that regret is minimized instead of maximizing utility. As we show in the following, it is therefore straightforward to define mother logit models that are equivalent to the RRM ones.

Let $r(a, a'|k)_t = \beta_t(x(a'|k)_t - x(a|k)_t)$ denote the regret difference between links a and a' associated with attribute t, and similarly let $v(a, a'|k)_t = \beta_t(x(a|k)_t - x(a'|k)_t)$ denote the utility difference. Chorus (2014) uses $\ln(\lambda_t + e^{r(a,a'|k)_t})$ as an approximation of $\max\{0, r(a, a'|k)_t\}$. Consequently, we use $-\ln(\lambda_t + e^{-v(a,a'|k)_t})$ as an approximation of $\min\{0, v(a, a'|k)_t\}$. As an illustration, we present in Fig. 2 a graph showing regret and utility attribute differences as functions of $\lambda_t \in [0, 1]$.

Based on this approximation we define three Equivalent Mother Logit (EML) models so that we have an EML equivalent to each of the RRM models (GRRM, ERRM and ARRM). Henceforth, RUM models refer to models proposed in Fosgerau et al. (2013). The deterministic utilities of the three EML models are

$$v^{\text{EML}}(a|k) = \sum_{a'\in A(k)} \sum_{t} \ln\left(\frac{1}{\lambda_t + e^{-v(a,a'|k)_t}}\right), \quad \forall\, k \in \mathcal{A},\ a \in A(k), \tag{14}$$

$$v^{\text{EEML}}(a|k) = \sum_{a'\in A(k)} \sum_{t} \ln\left(\frac{1}{\lambda_t + e^{-(v(a,a'|k)_t + \delta_t x(a'|k)_t)}}\right), \quad \forall\, k \in \mathcal{A},\ a \in A(k), \tag{15}$$

and

$$v^{\text{AEML}}(a|k) = \frac{1}{|A(k)|} \sum_{a'\in A(k)} \sum_{t} \ln\left(\frac{1}{\lambda_t + e^{-(v(a,a'|k)_t + \delta_t x(a'|k)_t)}}\right), \quad \forall\, k \in \mathcal{A},\ a \in A(k). \tag{16}$$

We introduce the following propositions to show the equivalence between the RRM and EML models. We denote the parameters of the EML models by $(\lambda^{\text{EML}}, \beta^{\text{EML}}, \delta^{\text{EML}})$, and the parameters of the RRM models by $(\lambda^{\text{RRM}}, \beta^{\text{RRM}}, \delta^{\text{RRM}})$.

Proposition 1. *If $\lambda^{EML} = \lambda^{RRM}$ and $\beta^{EML} = \beta^{RRM}$, then given any two links $a, k \in \widetilde{\mathcal{A}}$, $a \in A(k)$, we have*

$$P^{EML}(a|k) = P^{GRRM}(a|k),$$

where $P^{EML}(a|k)$, $P^{GRRM}(a|k)$ are the probabilities of choosing link a conditional to k given by the EML and GRRM models, respectively.

Proof. If $\lambda^{\text{EML}} = \lambda^{\text{RRM}}$ and $\beta^{\text{EML}} = \beta^{\text{RRM}}$, then $v(a, a'|k) = -r(a, a'|k),\ \forall\, a, a' \in A(k),\ k \in \mathcal{A}$. So we have

$$v^{\text{EML}}(a|k) = -\sum_{a'\in A(k)} \sum_{t} \ln(\lambda_t + e^{r(a,a'|k)_t}) = -r^{\text{GRRM}}(a|k),$$

According to (8), the entries of matrix M^d becomes

$$M^d_{ka} = \xi(a|k) e^{\frac{1}{\mu} v^{\text{EML}}(a|k)}, \quad \forall\, k,\ a \in \widetilde{\mathcal{A}}.$$

We note that the EML model is based on the RUM framework and M^d is defined as in Fosgerau et al. (2013). Then, z^d is a solution to the system of linear equations $(I - M^d)z^d = b$, so we have

$$z_k^d = e^{-\frac{1}{\mu}R^d(k)} = e^{\frac{1}{\mu}V^d(k)}, \quad \forall\, k \in \widetilde{A}.$$

The probability of choosing a link a given link k can hence be written as

$$P^{\mathrm{GRRM}}(a|k) = \xi(a|k)e^{-\frac{1}{\mu}(r(a|k)+R^d(a)-R^d(k))} = \xi(a|k)e^{\frac{1}{\mu}(v(a|k)+V^d(a)-V^d(k))} = P^{\mathrm{EML}}(a|k). \square$$

Proposition 2. *If $\lambda^{EML} = \lambda^{RRM}$, $\beta^{EML} = \beta^{RRM}$ and $\delta^{EML} = -\delta^{RRM}$ then*

$$P^{ERRM}(a|k) = P^{EEML}(a|k), \quad P^{\mathrm{ARRM}}(a|k) = P^{\mathrm{AEML}}(a|k), \quad \forall\, a \in A(k),\ k \in \mathcal{A},$$

where $P^{ERRM}(a|k)$, $P^{ARRM}(a|k)$, $P^{EEML}(a|k)$, $P^{AEML}(a|k)$ are the probabilities of choosing link a conditional to k given by the ERRM, ARRM, EEML and AEML models, respectively.

Proof. This is trivially verified, similarly to Proposition 1.□.

6. Maximum likelihood estimation

There are different ways of estimating a dynamic discrete choice model (see for instance Aguirregabiria and Mira, 2010). Similar to Fosgerau et al. (2013) and Mai et al. (2015a) we use the nested fixed point algorithm proposed by Rust (1987). This algorithm combines an outer iterative nonlinear optimization algorithm for searching over the parameter space with an inner algorithm for solving the expected minimum regrets or the expected maximum utilities (or the value functions). The value functions can be solved quickly using the system of linear equations in (10). The log-likelihood function for the EML models can be derived as in Fosgerau et al. (2013). We therefore turn our attention to the definition of the log-likelihood function as well as its derivatives for the RRM models.

The log-likelihood function defined for N observations $\sigma_1, \ldots, \sigma_N$ with respect to the vector of model parameters β is

$$LL(\beta) = \sum_{n=1}^{N} \ln P(\sigma_n) = \frac{1}{\mu}\sum_{n=1}^{N}\sum_{i=0}^{J_n-1}(R(k_0^n) - r(\sigma_n)).$$

For notational simplicity we omit the superscript d indicating the destinations but note that the choice probabilities $P(\sigma_n)$ and expected minimum regrets $R(k_0^n)$ depend on the destination of path σ_n. Efficient nonlinear techniques for the problem require analytical derivatives of the log-likelihood function. We therefore derive the gradient of $LL(\beta)$ with respect to a parameter β_i as

$$\frac{\partial LL(\beta)}{\partial \beta_i} = \frac{1}{\mu}\sum_{n=1}^{N}\sum_{i=0}^{J_n-1}\left(\frac{\partial R(k_0^n)}{\partial \beta_i} - \frac{\partial r(\sigma_n)}{\partial \beta_i}\right),$$

which requires the derivatives of $R(k_0^n)$. Taking the first derivative of (10), we obtain

$$\frac{\partial z}{\partial \beta_i} = (I - M)^{-1}\frac{\partial M}{\partial \beta_i}z, \quad \text{and using } \frac{\partial R(k)}{\partial \beta_i} = -\mu\frac{\partial z_k}{z\partial \beta_i}. \tag{17}$$

The gradients of the regret value function $R(k)$, $k \in \widetilde{A}$, can be quickly computed using the system of linear equations (17). The value of $r(\sigma)$ for a given path σ is nonlinear in parameters, so that $\frac{\partial r(\sigma)}{\partial \beta_i}$ has a complicated form but is fast to compute. We note that from (11), (12) and (13) the regret-based models have three vector of parameters to be estimated i.e. λ, β and δ. The GRRM model requires $0 \le \lambda_t \le 1$ for all attributes t. This implies that the MLE becomes a constrained optimization problem as in the following

$$\max_{\substack{\lambda,\beta,\delta \\ 0\le\lambda_t\le 1,\ \forall t}} LL(\lambda, \beta, \delta).$$

We use an interior point algorithm with BFGS to solve this constrained problem. The code is implemented in MATLAB (available upon request) and we use the function *fmincon* for solving the problem.

It is important to note that the RRM and EML models require comparisons between each alternative with every other one, attribute per attribute. This makes the estimation of the RRM and EML models more costly, compared to the classical RUM models in Fosgerau et al. (2013). In order to simplify the estimation, we use the decomposition (DeC) method (Mai et al., 2016) designed to speed up the estimation of RL models. This approach requires to solve one linear system only when computing the log-likelihood function, instead of solving one system of linear equations for each observed destination or origin-destination pair. The drawback of this method is that it is not compatible with the link size (LS) attribute (for instance Fosgerau et al., 2013) and the NRL model (Mai et al., 2015a).

7. Numerical results

In order to have comparable numerical results with previous studies, we use the same data as Fosgerau et al. (2013) (also used in Frejinger and Bierlaire, 2007; Mai et al., 2015b, 2015a), collected in the city of Borlänge, Sweden. This network is composed of 3077 nodes and 7459 links and it is uncongested so travel times are assumed static and deterministic. There are 1832 observations containing 466 destinations, 1420 different origin-destination (OD) pairs and more than 37,000 link choices. Moreover, as we explained in the following we specify the link regret functions using the same attributes as Fosgerau et al. (2013) and Mai et al. (2015a).

7.1. Model specifications

Four attributes are included in the regret function: travel time $TT(a)$ of link a, left turn $LT(a|k)$ that equals one if the turn angle from k to a is larger than 40 degrees and less than 177 degrees, link constant $LC(a)$ that equals one except the dummy link which equals zero and U-turn $UT(a|k)$ that equals one if the turn angle is larger than 177.

For the sake of comparison, we report the estimation and prediction results for the RUM-based RL (Fosgerau et al., 2013) and NRL (Mai et al., 2015a) models, their deterministic utility specifications are

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

where LS is the link size attribute (for a detailed description see Fosgerau et al., 2013). It has been computed using a linear-in-parameter formulation of the aforementioned attributes using parameters $\widetilde{\beta}_{TT} = -2.5, \widetilde{\beta}_{LT} = -1, \widetilde{\beta}_{LC} = 0.4, \widetilde{\beta}_{UT} = -4$. This attribute can be considered as a correction for the utilities in order to relax the IIA property from the RL model, similar to the path size (PS) attribute (Ben-Akiva and Bierlaire, 1999). We note that when the LS or PS is included, the corresponding models are no longer RUM models, but mother logit ones. The NRL model has the same instantaneous utility but the IIA is relaxed by allowing the random terms to have link specific scale parameters.

The regret specifications for the RRM models can be defined based on (11), (12) and (13), respectively, using the same four attributes as the RUM models. As mentioned earlier, the DeC method (Mai et al., 2016) is used to estimate the resulting expensive estimation problems in a reasonable time, and the LS attribute and the NRL model are not compatible with the DeC method. We therefore only estimate the RRM and EML models based on the RL with the aforementioned attributes.

There is an important difference related to the LC attribute. In the RUM model, the rationale behind using $LC(a)$ in the instantaneous utilities is to penalize paths with many crossings (links). In the regret context, the link constant equals one except for the destination which equals zero. So this attribute cancels out when comparing two outgoing links except when comparing a link in $\mathcal{A}$ with destination d. More precisely, for each link $k \in \mathcal{A}$, the regret for the ERRM model can be expressed as

$$r^{\mathrm{ERRM}}(a|k) = \sum_{t,t\neq LC}\sum_{a'\in A(k)}\left(\lambda_t + e^{\beta_t(x(a'|k)_t - x(a|k)_t)+\delta_t x(a'|k)_t}\right) + \psi(a|k)_{LC}, \tag{18}$$

where

$$\psi(a|k)_{LC} = \begin{cases}\sum_{a'\in A(k)}\ln(\lambda_{LC} + e^{\delta_{LC}}), & \forall\, a\in A(k),\ a\neq d,\ d\notin A(k)\\ \sum_{a'\in A(k),\, a'\neq d}\ln(\lambda_{LC} + e^{\delta_{LC}}) + \ln(\lambda_{LC} + e^{-\beta_{LC}+\delta_{LC}}), & \text{if } d\in A(k),\ a\neq d\\ \sum_{a'\in A(k)}\ln(\lambda_{LC} + e^{\beta_{LC}+\delta_{LC}}), & \text{if } a = d.\end{cases} \tag{19}$$

Eqs. (18) and (19)indicate that the value of β_{LC} only affects the regret $r^{\mathrm{ERRM}}(a|k)$ if link k connects directly to d. The other RRM-based regrets and EML-based utilities can be written in a similar way. Hence, the link constant plays a different role in the RRM/EML models than in the RUM models; it is an attraction factor at the destination. Such a factor is actually important for the RRM and EML models to ensure that the probability of choosing the destination link (once arriving at the destination) is close to one. Such an attraction attribute is not needed (and does not affect the probabilities) in the RUM models since the instantaneous utilities are negative except for the destination that is zero. In order to make the distinction clear between these attributes, we call it *destination constant* (DC) in the RRM and EML models. Accordingly, the regrets for the three RRM models are

$$\begin{aligned} r^{\mathrm{ERRM}}(a|k) = \sum_{a'\in A(k)} \Bigg\{ & \left(\ln\lambda_{TT} + e^{\beta_{TT}(TT(a') - TT(a)) + \delta_{TT}TT(a')}\right)\\ & + \ln\left(\lambda_{LT} + e^{\beta_{LT}(LT(a'|k) - LT(a|k)) + \delta_{LT}LT(a'|k)}\right)\\ & + \ln\left(\lambda_{DC} + e^{\beta_{DC}(DC(a') - DC(a)) + \delta_{DC}DC(a')}\right)\\ & + \ln\left(\lambda_{UT} + e^{\beta_{UT}(UT(a'|k) - UT(a|k)) + \delta_{UT}UT(a'|k)}\right)\Bigg\}\\ & \forall\, k\in\mathcal{A},\quad a\in A(k), \end{aligned}$$

$$r^{\mathrm{ARRM}}(a|k) = \frac{1}{|A(k)|}r^{\mathrm{ERRM}}(a|k), \quad \forall\, k\in\mathcal{A},\quad a\in A(k),$$

$$r^{\mathrm{GRRM}}(a|k) = \sum_{a' \in A(k)} \left\{ \ln\left(\lambda_{TT} + e^{\beta_{TT}(TT(a')-TT(a))}\right) + \ln\left(\lambda_{LT} + e^{\beta_{LT}(LT(a'|k)-LT(a|k)}\right) + \ln\left(\lambda_{DC} + e^{\beta_{DC}(DC(a')-DC(a))}\right) + \ln\left(\lambda_{UT} + e^{\beta_{UT}(UT(a'|k)-UT(a|k))}\right) \right\}, \quad \forall\, k \in \mathcal{A},\ a \in A(k).$$

Similarly, the utilities for the EML models are

Table 1
Estimation results.

Parameters	GRRM/EML	ARRM/AEML	ERRM/EEML
$\widehat{\beta}_{TT}$	−0.15	−1.92	−0.37
Rob. Std. Err.	0.01	0.21	0.09
Rob. *t*-test(0)	−11.46	−8.98	−4.05
$\widehat{\beta}_{LT}$	−0.34	−1.8	−0.31
Rob. Std. Err.	0.02	0.41	0.08
Rob. *t*-test(0)	−15.36	−4.43	−3.84
$\widehat{\beta}_{UT}$	−5.89	−7.32	−5.32
Rob. Std. Err.	0.57	65.32	1.87
Rob. *t*-test(0)	−10.32	−0.11	−2.85
$\widehat{\beta}_{DC}$	12.92	99.99	23.18
Rob. Std. Err.	1.66	36.03	3.79
Rob. *t*-test(0)	7.77	2.77	6.11
$\widehat{\delta}_{TT}$	–	3.75(−3.75)	1.22(−3.75)
Rob. Std. Err.	–	0.46	0.21
Rob. *t*-test(0)	–	8.18(−8.18)	5.69(−5.69)
$\widehat{\delta}_{LT}$	–	0.12(−0.12)	0.09(−0.09)
Rob. Std. Err.	–	0.7	0.1
Rob. *t*-test(0)	–	0.17(−0.17)	0.89(−0.89)
$\widehat{\delta}_{UT}$	–	7.16(−7.16)	4.75(−4.75)
Rob. Std. Err.	–	53.66	1.31
Rob. *t*-test(0)	–	0.13(−0.13)	3.62(−3.62)
$\widehat{\delta}_{DC}$	–	−7.16(7.16)	−1.44(1.44)
Rob. Std. Err.	–	63.49	1.73
Rob.*t*-test(0)	–	−0.11(0.11)	−0.84(0.84)
$\widehat{\lambda}_{TT}$	8.13E−06	0.37	1.00
Rob. Std. Err.	–	0.31	–
Rob. *t*-test(0)	–	1.2	–
$\widehat{\lambda}_{LT}$	7.26E−06	1.00	8.29E−05
Rob. Std. Err.	–	–	–
Rob. *t*-test(0)	–	–	–
$\widehat{\lambda}_{UT}$	0.76	0.01	1.04E−04
Rob. Std. Err.	0.04	–	–
Rob. *t*-test(0)	17.86	–	–
$\widehat{\lambda}_{DC}$	0.46	0.58	0.48
Rob. Std. Err.	0.02	37.06	0.81
Rob. *t*-test(0)	21.62	0.02	0.59

Table 2
Final log-likelihood values.

Models	# parameters	Final log-likelihood values
RL	4	−6303.9
RL-LS	5	−6045.6
NRL	7	−6187.9
NRL-LS	8	−5952.0
GRRM/EML	8	−7931.6
ARRM/AEML	12	−5661.6
ERRM/EEML	12	−5500.4

$$v^{\text{EEML}}(a|k) = -\sum_{a'\in A(k)} \left\{ \ln\left(\lambda_{TT} + e^{-\beta_{TT}(TT(a)-TT(a'))-\delta_{TT}TT(a')}\right) + \ln\left(\lambda_{LT} + e^{-\beta_{LT}(LT(a|k)-LT(a'|k))-\delta_{LT}LT(a'|k)}\right) + \ln\left(\lambda_{DC} + e^{-\beta_{DC}(DC(a)-DC(a'))-\delta_{DC}DC(a')}\right) + \ln\left(\lambda_{UT} + e^{-\beta_{UT}(UT(a|k)-UT(a'|k))-\delta_{UT}UT(a'|k)}\right) \right\} \quad \forall\, k \in \mathcal{A},\quad a \in A(k),$$

$$v^{\text{AEML}}(a|k) = \frac{1}{|A(k)|} r^{\text{EEML}}(a|k), \quad \forall\, k \in \mathcal{A},\ a \in A(k),$$

$$v^{\text{EML}}(a|k) = -\sum_{a'\in A(k)} \left\{ \ln\left(\lambda_{TT} + e^{\beta_{TT}(TT(a)-TT(a'))}\right) + \ln\left(\lambda_{LT} + e^{-\beta_{LT}(LT(a|k)-LT(a'|k)}\right) + \ln\left(\lambda_{DC} + e^{-\beta_{DC}(DC(a)-DC(a'))}\right) + \ln\left(\lambda_{UT} + e^{-\beta_{UT}(UT(a|k)-UT(a'|k))}\right) \right\}, \quad \forall\, k \in \mathcal{A},\ a \in A(k).$$

7.2. Estimation results

The estimation results for the three RRM models and three EML models are presented in Table 1. We report the estimates for the three pairs of models GRRM/EML, ARRM/AEML, and ERRM/EEML. For each row we report one value if the corresponding estimates are identical, otherwise we report the estimates for the EML models in the parentheses. Indeed, these results are consistent with Propositions 1 and 2.

For all the models, the $\widehat{\beta}$ are significantly different from zero except for the parameter associated with u-turns in the ARRM/AEML model. Moreover, they are, as expected, negative for travel time, left turns and u-turns. Based on the discussion in the previous section, we expect $\widehat{\beta}_{DC}$ to be positive and with large magnitudes so that $P(d|k)$ are close to one.

We now turn our attention to the δ estimates. The ERRM/EEML and ARRM/AEML models include δ so that they can flexibly capture the impact of non-chosen alternatives in the utilities or regrets. If $\widehat{\delta}_t > 0$, the impact of non-chosen alternatives is larger than if $\widehat{\delta}_t < 0$. For the RRM models, the estimation results show that $\widehat{\delta}_t$ are either not significantly different from zero, or they are significant and positive ($\widehat{\delta}_{TT}$ in the ARRM and $\widehat{\delta}_{TT}$, $\widehat{\delta}_{UT}$ in the ERRM model). It means that the impact of the non-chosen alternatives in the ERRM is larger than the GRRM model in terms of travel time and u-turns. On the contrary, for the EML models, δ estimates have the opposite signs of the ones given by the corresponding RRM models, meaning that they capture the opposite impact of those from the RRM models. We recall that if $\lambda_t = 0$ the regret/competitive utility associated with attribute t is linear-in-parameters and if $\lambda_t = 1$ the regret becomes the original RRM model proposed by Chorus (2012).

The last four rows of Table 1 show the λ estimates. Note that we do not provide standard errors and t-tests for the estimates that

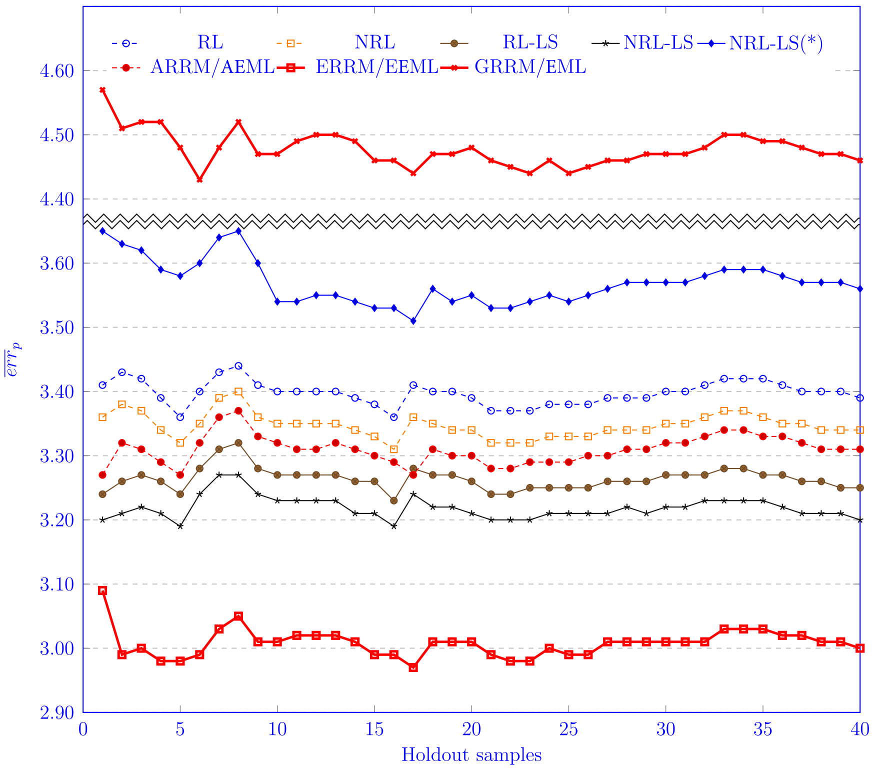

Fig. 3. Average of the test error values over holdout samples.

are on the bounds (close to 0 or 1) since the corresponding gradient components are not equal to zero. For the GRRM/EML model, the $\widehat{\lambda}_t$ are only significantly different from zero for the parameters associated with u-turns and destination constant. The others are close to zero (on the bound). However, for the ERRM/EEML or ARRM/AEML models, the $\widehat{\lambda}_t$ are either on the bounds ($\widehat{\lambda}_{LT}$, $\widehat{\lambda}_{UT}$ for the ARRM/AEML and $\widehat{\lambda}_{TT}$, $\widehat{\lambda}_{LT}$, $\widehat{\lambda}_{UT}$ for the ERRM/EEML), or not significantly different from zero ($\widehat{\lambda}_{TT}$, $\widehat{\lambda}_{DC}$ for the ARRM/AEML and $\widehat{\lambda}_{DC}$ for the ERRM/EEML). Chorus (2014) provides more detailed discussions on how the regrets change when parameters λ vary in the interval $[0, 1]$. These discussions can be applied to the EML models as well.

We report the final log-likelihood values in Table 2. The likelihood ratio test cannot be used to statistically compare the in-sample fit between RUM, RRM and EML models. Among the RUM models, the NRL ones have significantly better fit than the RL models. The models with the LS attribute are better than the ones without. Among the RRM and EML models, the ERRM/EEML perform better than the GRRM/EML. Moreover, since the RL-RUM model is a restricted version of the ARRM/AEML models, the results show that the ARRM and AEML have significantly better fit than the RL-RUM. Finally, we note that the ERRM/EEML have the highest and the GRRM/EML have the lowest final log-likelihood values.

Before discussing the out-of-sample fit of these models in the following section, we make some remarks about the computational time for the estimation. The RRM and EML models are more expensive to estimate than the RUM models due to the cost associated with computing the link regrets/utilities. Indeed, at each choice stage of the RRM- or EML models, we compare each alternative with every other alternative in the choice set. We use a non parallelized MATLAB code running under a core i5, Intel(R) 3.20 GHz machine with a x64-based processor for the maximum likelihood estimation.

For the RRM and EML models, given a vector of parameters, we need approximately 45 s to compute the link regrets/utilities. It

Table 3
Average of out-of-sample error values.

Models	Error values
RL	3.39
RL-LS	3.25
NRL	3.36
NRL-LS	3.20
NRL-LS(*)	3.56
GRRM/EML	4.46
ARRM/AEML	3.31
ERRM/EEML	3.00

takes less than 1 s for the RUM models as we can use matrix operations to compute linear-in-parameters utilities. We would therefore need approximately 4 h to compute the log-likelihood function of RRM/EML without the DeC approach, as many linear systems have to be solved. This computation only requires a few minutes with the DeC method because it allows to evaluate the log-likelihood function solving a single system of linear equations. The computation of the log-likelihood function for the RUM models takes approximately 5 min without DeC, and it takes less than 10 s with.

The estimation of the RRM and EML models is also expensive due to the nonlinearities in the utilities/regrets. The nonlinear optimization algorithm needs approximately 30 iterations to converge for the RUM models while the RRM/EML models require 300–800 iterations. So, with the DeC method, the RRM and EML models can be estimated in approximately one day and without the DeC method, the estimation would take several weeks. This explains why the LS attribute and the NRL model are too expensive to use with the RRM/EML models.

7.3. Prediction results

In this section we report results from a cross-validation study. The objective is to compare the out-of-sample fit of the models which is useful to detect overfitting and assess prediction performance.

We repeatedly divide the sample into two sets by uniformly drawing observations: we use one set containing 80% of the observations for estimation and the other (20%) as a holdout sample to evaluate the predicted probabilities. We generate 40 holdout samples of the same size by reshuffling the real sample and use the log-likelihood loss as the loss function to evaluate the prediction performance.

For each holdout sample i, $0 \le i \le 40$ we estimate the parameters β^i, δ^i and λ^i of the corresponding training sample and use these parameters to compute the test errors err_i

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

where PS_i is the size of the prediction sample i. We then compute the average of err_i over samples in order to have unconditional test error values

$$\overline{err}_p = \frac{1}{p}\sum_{i=1}^{p} err_i \quad \forall\ 1 \le p \le 40. \tag{20}$$

For comparison we also report the prediction performance of the four RUM models.

We apply the cross-validation study for all the RUM-based RL and NRL models, and the RRM- and EML-based RL models proposed in this paper. Moreover, as mentioned earlier, if we specify $\lambda = 0$ and $\beta = \delta$ (for EML models) or $\beta = -\delta$ (for RRM models), then the ERRM and EEML are equivalent to linear-in-parameter RUM models in which the utilities $v(a|k)$, $a \in A(k)$, are multiplied with $|A(k)|$. For the sake of comparison, we also apply the cross-validation for the RUM-based NRL-LS model, i.e., the best RUM model in prediction (for instance Mai et al., 2015a) where the link utilities are scaled as $|A(k)|v(a|k)$, $a \in A(k)$. We denote this RUM model by NRL-LS(*).

The values of $\overline{err}_p$, $1 \le p \le 40$ are shown in the graph in Fig. 3 . We also report the averages of the errors over all the holdout samples in Table 3. In line with Propositions 1 and 2, the prediction results given by the GRRM, ERRM, ARRM are identical to the EML, EEML and AEML models, respectively. As expected, the value of $\overline{err}_p$ for each model stabilizes as p increases. The results show that the ERRM/EEML models perform the best (lowest value of the loss function). The GRRM/EML models have by far the worst out-of-sample fit. Interestingly, we observe overfitting in the ARRM/AEML models, as they have final log-likelihood values (in-sample fit) that are almost 300 units better than the best RUM model (NRL-LS). However, the prediction performance is worse than both NRL-LS and RL-LS. Moreover, the NRL-LS(*) model performs worse than all the other models except the GRRM/EML. Since the ERRM/EEML perform better than the RUM-based RL, this remark indicates the important role of the non-chosen alternatives in the RRM and EML models.

8. Conclusion

In this paper, we have compared the in-sample and out-of-sample fit of random maximization, random regret minimization and equivalent mother recursive logit route choice models. We adapted the GRRM model proposed by Chorus (2014) and proposed two variants: ARRM and ERRM models. We showed that we can define mother logit models that, under some conditions, are equivalent to the RRM ones. We provided numerical results and a cross-validation study using real data and a network of more than 3000 nodes and 7000 links. The cross-validation results showed that the ERRM/EEML have the best in-sample and out-of-sample fit and that the GRRM/EML have the worst fit. These results indicate that for the application at hand, it is important to include the additional terms associated with non-chosen alternatives in the regret/utility attribute differences. We note that other specifications of the RRM models have been recently proposed (see for instance van Cranenburgh et al., 2015), which could be interesting to investigate.

It is important to emphasize that the estimation and application of the RRM and EML models are more time consuming, compared to the RUM ones. In addition, the interpretation of the parameter estimates is less straightforward.

Acknowledgements

This research was funded by the National Sciences and Engineering Research Council of Canada, discovery grant 435678-2013. A preliminary version of this paper was presented at the 2015 International Choice Modeling Conference and we are grateful for valuable comments from the participants and discussions with Caspar Chorus. We would also like to thank Hani Mahamassani for interesting discussions and for encouraging us to work on the equivalence between the RRM and mother logit models. The paper has also been improved thanks to the comments of the two anonymous reviewers.

References

Aguirregabiria, Victor, Mira, Pedro, 2010. Dynamic discrete choice structural models: a survey. J. Econ. 156 (1), 38–67.
Ben-Akiva, Moshe, 1973. The Structure of Travel Demand Models. (Ph.D. thesis), MIT.
Ben-Akiva, Moshe, Bierlaire, Michel, 1999. Discrete choice methods and their applications to short term travel decisions. In: Handbook of Transportation Science, Springer. pp. 5–33.
Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In: Lesort, J.B. (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France.
Chorus, Caspar G., 2010. A new model of random regret minimization. Eur. J. Transp. Infrastruct. Res. 2 (10), 181–196.
Chorus, Caspar G., 2012. Random regret minimization: an overview of model properties and empirical evidence. Transp. Rev. 32 (1), 75–92.
Chorus, Caspar G., 2014. A generalized random regret minimization model. Transp. Res. Part B 68 (1), 224–238.
Daly, Andrew, Bierlaire, Michel, 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., 2013. A link based network route choice model with unrestricted choice set. Transp. Res. Part B 56, 70–80.
Frejinger, E., Bierlaire, M., Ben-Akiva, M., 2009. Sampling of alternatives for route choice modeling. Transp. Res. Part B 43 (10), 984–994.
Frejinger, Emma, Bierlaire, Michel, 2007. Capturing correlation with subnetworks in route choice models. Transp. Res. Part B 41 (3), 363–378.
Horowitz, Joel L., Louviere, Jordan J., 1995. What is the role of consideration sets in choice modeling? Int. J. Res. Mark. 12 (1), 39–54.
Kahneman, Daniel, Tversky, Amos, 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–291.
Lai, Xinjun, Bierlaire, Michel, 2015. Specification of the cross nested logit model with sampling of alternatives for route choice models. Transp. Res. Part B 80 (1), 220–234.
Loomes, Graham, Sugden, Robert, 1982. Regret theory: an alternative theory of rational choice under uncertainty. Econ. J. 92 (368), 805–824.
Mai, Tien, 2016. A method of integrating correlation structures for a generalized recursive route choice model. Transp. Res. Part B 93, 146–161.
Mai, Tien, Fosgerau, Mogens, Frejinger, Emma, 2015a. A nested recursive logit model for route choice analysis. Transp. Res. Part B 75 (1), 100–112.
Mai, Tien, Frejinger, Emma, Bastin, Fabian, 2015b. A misspecification test for logit based route choice models. Econ. Transp. 4 (4), 215–226.
Mai, Tien, Bastin, Fabian, Frejinger, Emma, 2016. A decomposition method for estimating recursive logit based route choice models. Eur. J. Transp. Logist., 1–23. http://dx.doi.org/10.1007/s13676-016-0102-3.
McFadden, Daniel, 2000. Disaggregate behavioral travel demand's RUM side. Travel Behav. Res., 17–63.
McFadden, Daniel, Tye, William B., Train, Kenneth, 1978. An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model. Transport. Res. Rec.: Forecasting Passenger Freight Travel 637, 39–46.
McFadden, Daniel L., 1984. Econometric analysis of qualitative response models. Handb. econometrics 2, 1395–1457.
McFadden, Daniel L., Train, Kenneth, 2000. Mixed MNL models for discrete response. J. Appl. Econ. 15 (5), 447–470.
Prato, Carlo Giacomo, 2014. Expanding the applicability of random regret minimization for route choice analysis. Transportation 41 (2), 351–375.
Rust, John, 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. Econometrica 55 (5), 999–1033.
van Cranenburgh, Sander, Guevara, Cristian Angelo, Chorus, Caspar G., 2015. New insights on random regret minimization models. Transp. Res. Part A 74 (1), 91–109.
Vovsha, P., Bekhor, S., 1998. Link-nested logit model of route choice overcoming route overlapping problem. Transp. Res. Rec. 1645, 133–142.