CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

Comparing Regret Minimization and Utility Maximization for Route Choice using the Recursive Logit Model

Tien Mai
Emma Frejinger
Fabien Bastin

June 2015

CIRRELT-2015-25

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Comparing Regret Minimization and Utility Maximization for Route Choice using the Recursive Logit Model

Tien Mai, Emma Frejinger*, Fabian Bastin

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Computer Science and Operations Research, Université de Montréal, P.O. Box 6128, Station Centre-Ville, Montréal, Canada H3C 3J7

Abstract. This paper focuses on the comparison of estimation and prediction results between the random utility maximization (RUM) and random regret minimization (RRM) frameworks for the route choice recursive logit (RL) model (Fosgerau et al., 2013). The RL model is originally based on the RUM framework. We propose different versions of the RL model based on the RRM framework, by adapting and extending the model proposed by Chorus (2014). We report estimation results and a cross-validation study for a real network with more than 3000 nodes and 7000 links. The cross-validation results show that one of the proposed extended version of the RRM-based model has the best out-of-sample fit. While this observation favours the RRM framework, we note that the RRM-based models are computationally more complex to estimate and apply than the RUM-based ones.

Keywords: Random regret, random utility, discrete choice.

Acknowledgement: This research was partly funded by the National Sciences and Engineering Research Council of Canada (NSERC), discovery grant 435678-2013.

---

* Corresponding author: Emma.Frejinger@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2015

1 Introduction

Discrete choice models are generally used for analyzing path choices in real networks based on revealed preference (RP) data. Following the discussion in Fosgerau et al. (2013) the route choice models in the literature can be grouped into three approaches. First, the classic approach corresponds to path logit (PL) models where choice sets of paths are generated and treated as the actual choice sets. The second approach, proposed by Frejinger et al. (2009), is based on the idea that the choice set can be sampled and the estimation can be consistent if the sampling protocols are added to the choice probabilities. Recently, Fosgerau et al. (2013) propose third approach, called the recursive logit (RL) model which can be consistently estimated based on RP data and used for prediction without sampling any choice sets of paths. An other extension of the RL model, the nested RL model, has been proposed by Mai et al. (2015) that allows to relax the independence of irrelevant alternatives (IIA) property. These models are based on the Random Utility Maximization (RUM) framework.

Recently, Chorus (2012, 2010, 2014) proposed the Random Regret Minimization (RRM) framework. It is based on a regret minimization-based decision rule postulating that when decision makers choose between alternatives, they try to avoid the situation where a non-chosen alternative outperforms a chosen one in terms of attributes. We base this paper the so-called Generalized Random Regret Minimization (GRRM) model proposed by Chorus (2014). In this paper we adapt and compare the estimation and prediction results of the RL model using RRM and RUM decision rules. Prato (2012) analyses the estimation results of path based models using the model proposed by Chorus (2010) in a route choice context. He focuses on the two well-known challenges associated with route choice modeling, namely, choice set generation and correlation. He finds that the RRM performs well on real data, but in an experimental setting, he finds that the parameter estimates of the RRM models have the wrong signs when irrelevant alternatives are included in the choice sets. The RL model is based on the universal choice set of all paths connecting an origin-destination pair. We investigate whether this issue occurs for the RL model using real data and we also analyze the models' out-of-sample fit.

This paper makes a number of contributions. First, we adapt and propose two specifications for random regret. The first model (called ERRM) extends the GRRM model by adding factors that allow to capture the impact of the non-chosen alternatives in a more flexible way and the second model (ARRM) modifies the first one by adding a normalization. We can prove that by specifying some parameters, the regret given by the ARRM model model has a linear-in-parameters form $R_{ni} = -\beta^T x_{ni}$ and the probability is equivalent to a RUM-based model with utility $V_{ni} = \beta^T x_{ni}$. This model therefore generalizes the RUM-based RL model.

Second, we show how the regret-based with RL model can be efficiently estimated. Third, we provide estimation and cross-validation results for a real network with over 3000 nodes and 7000 links. The estimation code for the RRM-based RL models is implemented in MATLAB and is freely available upon request.

The paper is structured as follows. In Section 2 we review the RUM- and RRM-based models. Section 3 proposes the RL model using the regret decision rules with with two different formulas for the regrets. In Section 4 we present the log-likelihood function and its gradient. Model specifications as well as estimation and cross-validation results are presented in Section 5, and finally, Section 6 concludes.

2 Random utility maximization and random regret minimization models

In the context of the RUM-based discrete choice models, we assume that an individual n associates a utility U_{ni} with an alternative i within a choice set C_n The utility consists of two additive parts $U_{ni} = V_{ni} + \epsilon_{ni}$: a deterministic V_{ni} part, observed by the modeller, and a random part ϵ_{ni}. Typically, a linear-in-parameters formula associated with a vector of attributes is used: $V_{ni} = \sum_t \beta_t x_{ni}(t)$, 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

$$i^* = \text{argmax}_{i \in C_n} \{U_{ni}\}. \tag{1}$$

The RRM-based models are based on the assumption that when decision makers choose between alternatives, they 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 that by definition features all attributes of all competing alternatives. The random regret RR_{in} can be written as the sum of a systematic part R_{ni} and a random error term ϵ_{ni},

$$RR_{in} = R_{in} + \mu\epsilon_{in} = \sum_{j \neq i, j \in C_n} \sum_t \ln\left(1 + e^{\beta_t(x_{jn}(t) - x_{in}(t))}\right) + \mu\epsilon_{in} \tag{2}$$

Contrary to the RUM-based models, a decision maker aims to minimize the random regret

$$\begin{aligned} i^* &= \text{argmin}_{i \in C_n} \left\{R_{in} + \epsilon_{in}\right\} \\ &= \text{argmax}_{i \in C_n} \left\{-R_{in} - \epsilon_{in}\right\}. \end{aligned}$$

Under the assumption that the random terms ϵ_{ni} are i.i.d extreme value type I, the choice probability is

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

Even though this is the logit model, the IIA property is relaxed since the regrets are not alternative specific. Chorus (2014) recently presented a generalization of the RRM model, called GRRM, where the random regret can be expressed as

$$RR_{in} = GR_{in} + \mu\epsilon_{in} = \sum_{j\neq i, j\in C_n} \sum_t \ln\left(\lambda_t + e^{\beta_t(x_{jn}(t)-x_{in}(t))}\right) + \mu\epsilon_{in} \tag{3}$$

When the parameter λ_t $\forall t$ are equal to 1 the GRRM model becomes the RRM model. If $\lambda = 0$ $\forall t$, we can express the regret as

$$GR_{in} = \sum_{j\neq i, j\in C_n} \sum_t \beta_t(x_{jn}(t) - x_{in}(t)) = \sum_{j\in C_n} \beta^T x_{jn} + |C_n|\beta^T x_{in}. \tag{4}$$

Since $\sum_{j\in C_n} \beta^T x_{jn}$ is equal for all $j \in C_n$, it does not affect the choice probabilities and the regret has a linear-in-parameter formulation (although different from the RUM-based model because of $|C_n|$. The term $\sum_{j\in C_n} \beta^T x_{jn}$ plays however a role in the logsum $\mathbb{E}\Big[\min_{i\in C_n}\Big\{GR_{ni} + \epsilon_{ni}\Big\}\Big]$. A disadvantage of the RRM or GRRM model, highlighted in Chorus (2012), is that its running times for computing the choice probabilities may suffer from combinatorial explosion when choice sets become very large and every alternative is compared with every other alternative in the choice set in terms of every attribute. The RL model is based on the universal choice (infinite size) but the choice set at each choice stage is small, just the outgoing links at a node. This is therefore not an issue for the RL model. The fact that the regret GR_{in} or RR_{in} is undefined when the choice set C_n is singleton is however an issue for the RL model because there can be only one outgoing link at a node in a transport network. We discuss how we deal with that in the following section.

3 Recursive logit with regret-based models

The RUM-based RL (RL-RUM) model formulates the path choice problem as a sequence of link choices, represented in a dynamic discrete choice framework. A utility is associated with each link pair in the network, it is the sum of a deterministic and a random term (independently and identically distributed, i.i.d., extreme value type I). Fosgerau et al. (2013) consider a linear-in-parameters formulation of the deterministic utility. A traveller maximizes 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 (value function). In the following we present the RRM-based RL model, the derivation is similar to Fosgerau et al. (2013) but the utilities and value functions are different since they are based on random regret minimization.

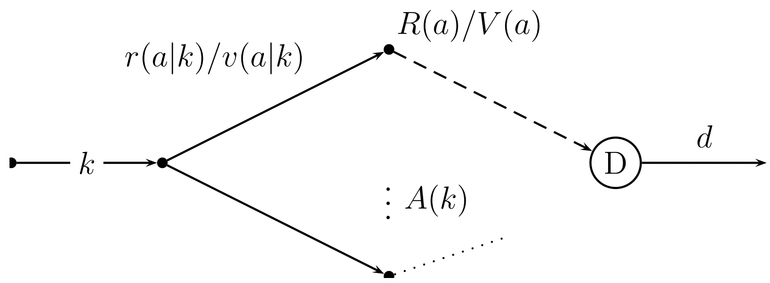

Figure 1: Illustration of notation

A directed connected graph (not assumed acyclic) $G = (A; \mathcal{V})$ is considered, where A and $\mathcal{V}$ are the set of links and nodes, respectively. For each link $k \in A$, 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 per destination that has no successors, that is, an absorbing state. The set of all links for a given destination is hence $\tilde{A} = A \cup \{d\}$. Given two links $a, k \in \tilde{A}$, the following instantaneous random regret is associated with action $a \in A(k)$

$$rr_n(a|k) = r_n(a|k) + \mu\epsilon_n(a) \tag{5}$$

where $r_n(a|k)$ is a regret associated with link a given link k and $\epsilon_n(a)$ are the random terms. At each current state k the traveler observes the realizations of the random terms $\epsilon_n(a)$, $a \in A(k)$. He/she then chooses link a that minimizes the sum of instantaneous random regret $rr_n(a|k)$ and expected downstream regret. The latter is defined as the expected minimum regret from state k to the destination is $R^d(k)$, which is recursively defined by Bellman's equation as

$$R^d(k) = \mathbb{E}\left[\min_{a\in A(k)} \left\{rr_n(a|k) + R^d(a) + \mu\epsilon_n(a)\right\}\right], \quad \forall k \in A. \tag{6}$$

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 $r_n(a|k) = r_n(a|k; \beta)$ where β is the parameters to be estimated. For notational simplicity we omit from now on β from $R(\cdot)$ and $r(\cdot)$. We define regret $r_n(a|k)$ based on the GRRM model (Chorus, 2014). It is important to consider that $A(k)$ may contain only one link. Existing random regret models would in this case assign a regret zero which would cause numerical issues for the RL model since it must be costly to travel. We therefore define the regret based on all outgoing links and the slightly modified GRRM is

$$r_n(a|k) = \sum_{a'\in A(k)} \sum_t \ln\left(\lambda_t + e^{\beta_t(x^n(a'|k)_t - x^n(a|k)_t)}\right), \quad \forall a \in A(k), \tag{7}$$

where $r_n(a|k)$ is a regret associated with link a given link k, $x^n(a|k)_t$ is attribute t associated with link a given k, λ_t and β_t are 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.

We also define a new formulation for regret that we call Extended Random Regret Minimization (ERRM) to capture the impact of non-chosen alternatives in a more flexible way

$$r_n(a|k) = \sum_{a' \in A(k)} \sum_t \ln\left(\lambda_t + e^{\beta_t \left(x^n(a'|k)_t - x^n(a|k)_t\right) + \delta_t x^n(a'|k)_t}\right), \quad \forall a \in A(k). \tag{8}$$

The difference lies in the term $\delta_t x^n(a'|k)_t$. 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. For the same reason, we omit an index for individual n but note that regrets $r(a|k)$ can be individual specific. Equation (6) can be written as

$$\begin{aligned} R^d(k) &= \mathbb{E}\left[-\max_{a \in A(k)} \left\{-r(a|k) - R^d(a) - \mu\epsilon(a|k)\right\}\right] \\ &= -\mathbb{E}\left[\max_{a \in A(k)} \left\{-r(a|k) - R^d(a) + \mu(-\epsilon(a|k))\right\}\right], \quad \forall k \in A. \end{aligned} \tag{9}$$

Following Chorus (2010) we assume that random terms $-\epsilon(a|k)$ are i.i.d. standard extreme value type I and the probability of choosing link a given state k is given by the MNL model

$$P^d(a|k) = \frac{\delta(a|k) e^{-\frac{1}{\mu}(r(a|k) + R^d(a))}}{\sum_{a' \in A(k)} e^{-\frac{1}{\mu}(r(a'|k) + R^d(a'))}}, \quad \forall a, k \in A. \tag{10}$$

Note that we include $\delta(a|k)$ that equals one if $a \in A(k)$ and zero otherwise so that the probability is defined for all $a, k \in \tilde{A}$ (we recall that $\tilde{A} = A \cup \{d\}$). The expected minimum regret in this case is given by the logsum

$$-\frac{1}{\mu} R^d(k) = \ln\left(\sum_{a' \in A(k)} e^{\frac{1}{\mu}(-r(a|k) - R(a))}\right) \tag{11}$$

and $R^d(d) = 0$ by assumption. Equation (11) can be written as

$$e^{-\frac{1}{\mu}R(k)} = \begin{cases} \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}(-r(a|k) - R(a))} & k \in A, k \neq d. \\ 0 & k = d. \end{cases} \tag{12}$$

We define a matrix M^d of size $|\tilde{A}| \times |\tilde{A}|$ and a vector z of size $|\tilde{A}|$ with entries

$$M_{ka} = \delta(a|k)e^{-\frac{1}{\mu}r(a|k)},\ z_k = e^{-\frac{1}{\mu}R(k)}, \forall k, a \in \tilde{A}. \tag{13}$$

so that we obtain a system of linear equation for computing z as follows

$$z = (I - M)^{-1}b. \tag{14}$$

b is a vector of size $|\tilde{A}|$ with zeros values for all states except the dummy link d. Using (10), the probability of choosing link a given a state k can be written as

$$P^d(a|k) = \delta(a|k)e^{-\frac{1}{\mu}(r(a|k)+R(a)-R(k))}, \quad \forall k, a \in A \tag{15}$$

and the probability of a path $\sigma = [k_0, \dots, k_I]$

$$P(\sigma) = e^{\frac{1}{\mu}R^d(k_0)} \prod_{i=0}^{I-1} e^{-\frac{1}{\mu}r(a|k)} = e^{\frac{1}{\mu}R^d(k_0)} e^{-\frac{1}{\mu}r(\sigma)}, \tag{16}$$

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

$$\frac{P(\sigma_2)}{P(\sigma_2)} = e^{\frac{1}{\mu}(r(\sigma_2)-r(\sigma_1))} \tag{17}$$

and it does not depend only on the attributes of links in paths σ_1, σ_2. Hence, the IIA property does not hold for the RRM-based models.

By specifying $\lambda_t = 0$ and $\delta_t = -\beta_t$, for all attributes t, the regret given by ERRM model becomes

$$r(a|k) = -|A(k)|\beta^T x(a|k) = -|A(k)|v(a|k). \tag{18}$$

Hence, if $v(a|k)$ is linear-in-parameters as in Fosgerau et al. (2013), the regret is also 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. We consider an alternative of the ERRM model where a normalization factor is used so that the regret is averaged over all the alternatives as

$$ar(a|k) = \frac{1}{|A(k)|} r(a|k), \quad \forall a \in A(k). \tag{19}$$

We refer to this model as Averaged Random Regret Minimization (ARRM). Accordingly, by specifying $\lambda_t = 0$ and $\delta_t = -\beta_t$ we obtain $ar(a|k) = -v(a|k)$. Based on (13) the entries of matrix M are

$$M_{ka} = \delta(a|k)e^{\frac{1}{\mu}v(a|k)}, \forall k, a \in \tilde{A}.$$

We refer to the definition of the matrix M in Fosgerau et al. (2013) and note that z is a solution to the system of linear equations $z = (I - M)^{-1}b$, therefore it is straightforward to show that

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

where $V(k)$ 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

$$\begin{aligned} P^d(a|k) &= \delta(a|k)e^{-\frac{1}{\mu}(r(a|k)+R(a)-R(k))} \\ &= \delta(a|k)e^{\frac{1}{\mu}(v(a|k)+V(a)-V(k))}. \end{aligned} \tag{20}$$

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

4 Maximum likelihood estimation

In this section we present the log-likelihood function and its gradient. The log-likelihood function defined for N observations $\sigma_1, \ldots, \sigma_N$ with respect to the model parameters β is

$$LL(\beta) = \sum_{n=1}^{N} \ln P(\sigma_n) = \frac{1}{\mu} \sum_{n=1}^{N} \sum_{i=0}^{I_n-1} \left(R(k_0^n) - r(\sigma_n) \right) \tag{21}$$

The gradient of $LL(\beta)$ with respect to a parameter β_i is

$$\frac{\partial LL(\beta)}{\partial \beta_i} = \frac{1}{\mu} \sum_{n=1}^{N} \sum_{i=0}^{I_n-1} \left(\frac{\partial R(k_0^n)}{\partial \beta_i} - \frac{\partial r(\sigma_n)}{\partial \beta_i} \right) \tag{22}$$

To derive $\frac{\partial R(k_0^n)}{\partial \beta_i}$ we differentiate (14) which yields

$$\frac{\partial z}{\partial \beta_i} = (I - M)^{-1} \frac{\partial M}{\partial \beta_i} z \tag{23}$$

and using

$$\frac{\partial R(k)}{\partial \beta_i} = -\mu \frac{\partial z_k}{z \partial \beta_i}$$

The gradient of the regret value function $R(k)$, $k \in \tilde{A}$ can be efficiently computed thanks to the system of linear equations (23). We note that the value of $\frac{\partial M}{\partial \beta_i}$

and $\frac{\partial r(\sigma)}{\partial \beta_i}$ are computed using the derivatives of the regret functions which can be derived analytically as

$$\frac{\partial r(a|k)}{\partial \beta_i} = \sum_{a' \in A(k)} \sum_t \frac{\partial \phi_t(a, a'|k)}{\phi_t(a, a'|k) \partial \beta_i} \tag{24}$$

if the model ERRM is used, or

$$\frac{\partial r(a|k)}{\partial \beta_i} = \frac{1}{|A(k)|} \sum_{a' \in A(k)} \sum_t \frac{\partial \phi_t(a, a'|k)}{\phi_t(a, a'|k) \partial \beta_i} \tag{25}$$

for the ARRM model, where $\phi_t(a, a'|k) = \lambda_t + e^{\beta_t \left(x^n(a'|k)_t - x^n(a|k)_t \right) + \delta_t x^n(a'|k)_t}$. We note that the GRRM model requires $0 \leq \lambda_t \leq 1$ for all attributes t. This implies that the MLE becomes a constrained optimization problem. We use the interior point algorithm with BFGS to solve this constrained problem.

5 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., 2014, 2015), 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.

5.1 Model specifications

Four attributes are included in the regret function: link travel time $TT(a)$ of action a, number of 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.

The regret for the ERRM and ARRM model specifications are

$$
\begin{aligned}
r^{\mathrm{ERRM}}(a|k) = \sum_{a' \in A(k)} \Big\{ & \ln(\lambda_{TT} + e^{\beta_{TT}(TT(a')-TT(a))+\delta_{TT}TT(a')}) \\
& + \ln(\lambda_{LT} + e^{\beta_{LT}(LT(a')-LT(a))+\delta_{LT}LT(a')}) \\
& + \ln(\lambda_{LC} + e^{\beta_{LC}(LC(a')-LC(a))+\delta_{LC}LC(a')}) \\
& + \ln(\lambda_{UT} + e^{\beta_{UT}(UT(a')-UT(a))+\delta_{UT}UT(a')}) \Big\}
\end{aligned}
$$

and

$$
r^{\mathrm{ARRM}}(a|k) = \frac{1}{|A(k)|} r^{\mathrm{ERRM}}(a|k) \tag{26}
$$

Moreover, the regret for the GRRM model are

$$
\begin{aligned}
r^{\mathrm{GRRM}}(a|k) = \sum_{a' \in A(k)} \Big\{ & \ln(\lambda_{TT} + e^{\beta_{TT}(TT(a')-TT(a))}) \\
& + \ln(\lambda_{LT} + e^{\beta_{LT}(LT(a'|k)-LT(a|k))}) \\
& + \ln(\lambda_{LC} + e^{\beta_{LC}(LC(a')-LC(a))}) \\
& + \ln(\lambda_{UT} + e^{\beta_{UT}(UT(a'|k)-UT(a|k))}) \Big\}, \quad \text{if } |A(k)| > 1
\end{aligned} \tag{27}
$$

and

$$
r^{\mathrm{GRRM}}(a|k) = 0, \quad \text{if } |A(k)| = 1. \tag{28}
$$

5.2 Estimation results

The estimation results for the three models are presented in Table 2. The β parameter estimates have their expected signs and are significantly different from zero except for the parameter associated with u-turns in the ARRM model. If $\delta_t > 0$, the impact of non-chosen alternatives larger than if $\delta_t < 0$. The estimation results show that $\widehat{\delta_t}$ are either not significantly different from zero, or they are positive ($\widehat{\delta_{TT}}$ in the ARRM and $\widehat{\delta_{TT}}$, $\widehat{\delta_{UT}}$ in the ERRM model).

We report the final log-likelihood values in Table 1. For the sake of comparison, we report the values also for the RUM-based RL models Fosgerau et al. (2013), Mai et al. (2014). The differences in in-sample fit cannot be statistically compared with a likelihood ratio test because the models are not nested (except the RUM based ones where NRL-LS has significantly better fit than the other models). We note however that ERRM has the highest value.

Before discussing the out-of-sample fit of these models in the following section, we make some remarks about the computational time for estimation. The RRM-based models are more difficult to estimate than the RUM-based models due to the non-linearity in the utilities. The non-linear optimization algorithm needs approximate 30 iterations to converge for the RRM-based models while around 300 iterations are needed for the RRM-based ones.

	RL	RL-LS	NRL	NRL-LS	GRRM	ARRM	ERRM
# parameters	4	5	7	8	8	12	12
Final log-likelihood	-6303.9	-6045.6	-6187.9	-5952.0	-7931.6	-5661.6	-5500.4

Table 1: Final log-likelihood values

5.3 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.

Similar to Mai et al. (2015), the sample of observations is repeatedly divided into two sets by drawing observations at random with a fixed probability: one set contains 80% of the observations is used for estimation and the other (20%) is used as holdout samples to evaluate the predicted probabilities by applying the estimated model. We generate 40 holdout samples of the same size by reshuffling the real sample and use the the log-likelihood loss as the loss function to evaluate the prediction performance.

For each holdout sample i, $0 \leq i \leq 40$ we estimate the parameters $\widehat{\beta_i}$ of the corresponding training sample and these 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 the prediction sample i. We then compute the average of err_i values over samples in order to have unconditional test error values

$$\overline{err}_p = \frac{1}{p} \sum_{i=1}^{p} err_i \quad \forall 1 \leq p \leq 40. \tag{29}$$

For comparison we also report the predictions performances of the four RUM-based models.

The values of $\overline{err}_p$, $1 \leq p \leq 40$ are plotted in Figure 2 and Table 3 reports the average of the test error values over 40 samples. As expected, the value of $\overline{err}_p$

Parameters	GRRM	ARRM	ERRM
$\widehat{\beta}_{TT}$	-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	-1.80	-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
$\widehat{\beta}_{LC}$	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	1.22
Rob. Std. Err.		0.46	0.21
Rob. t-test(0)		8.18	5.69
$\widehat{\delta}_{LT}$		0.12	0.09
Rob. Std. Err.		0.70	0.10
Rob. t-test(0)		0.17	0.89
$\widehat{\delta}_{UT}$		7.16	4.75
Rob. Std. Err.		53.66	1.31
Rob. t-test(0)		0.13	3.62
$\widehat{\delta}_{LC}$		-7.16	-1.44
Rob. Std. Err.		63.49	1.73
Rob. t-test(0)		-0.11	-0.84
$\widehat{\lambda}_{TT}$	8.13e-6	0.37	1.00
Rob. Std. Err.	0.04	0.31	0.39
Rob. t-test(0)	1.80e-4	1.20	2.55
$\widehat{\lambda}_{LT}$	7.26e-6	1.00	8.29e-5
Rob. Std. Err.	0.06	0.45	1.50
Rob. t-test(0)	1.17e-4	2.24	5.53e-5
$\widehat{\lambda}_{UT}$	0.76	0.01	1.04e-4
Rob. Std. Err.	0.04	65.01	0.73
Rob. t-test(0)	17.86	0.00	1.44e-4
$\widehat{\lambda}_{LC}$	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: *Estimation results*

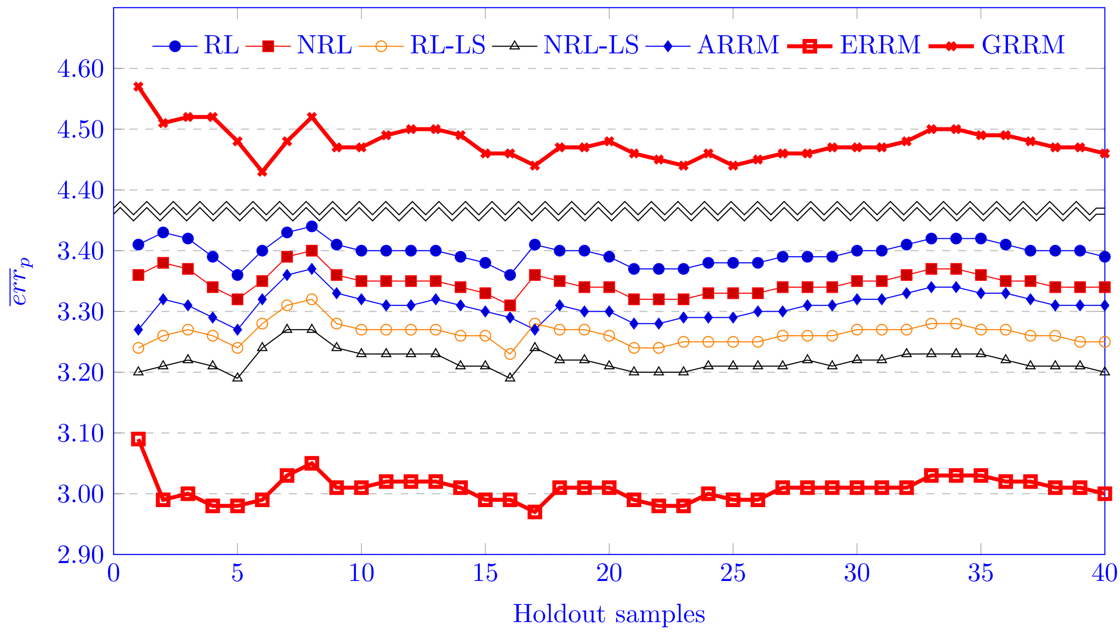

Figure 2: Average of the test error values over holdout samples

for each model stabilizes as p increases. The results show that the ERRM model performs best (lowest value of the loss function). The performance of the ERRM model is very different from GRRM that has the worst performance. Interestingly, the ARRM has a final log-likelihood value (in-sample fit) that is almost 300 units better than the best RUM-based model (NRL-LS) but the prediction performance is worse than both NRL-LS and RL-LS.

RL	RL-LS	NRL	NRL-LS	GRRM	ARRM	ERRM
3.39	3.25	3.36	3.20	4.46	3.31	3.00

Table 3: Average of test error values over 40 holdout samples

6 Conclusion

This paper compares estimation results and prediction performance between RUM and RRM based RL route choice models. We adapt from the GRRM model proposed by Chorus (2014) and propose two variants: ARRM and ERRM models. We derive the log-likelihood function as well as its gradient so that the RRM-based models. We provide numerical results and a cross-validation study using real data in a network with more 3000 nodes and 7000 links. The cross-validation results indicate that the ERRM model performs the best (it also has a higher final log-likelihood value in the estimation) and the performance of the GRRM model is much worse. These results indicate that RRM rule may be an interesting avenue for route choice modelling. It is however important to note that the estimation and application of the RRM based models are more complicated and time consuming than the RUM ones. Moreover, the interpretation of the parameter estimates are less straightforward.

Acknowledgment

This research was partly funded by the National Sciences and Engineering Research Council of Canada, discovery grant 435678-2013.

References

Chorus, C. Random regret minimization: An overview of model properties and empirical evidence. *Transport reviews*, 32(1):75–92, 2012.

Chorus, C. G. A new model of random regret minimization. *EJTIR*, 2(10), 2010.

Chorus, C. G. A generalized random regret minimization model. *Transportation Research Part B*, 68(1):224 – 238, 2014. ISSN 0191-2615.

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, 2013.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. Sampling of alternatives for route choice modeling. *Transportation Research Part B*, 43(10):984–994, 2009.

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

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

Prato, C. G. Estimating random regret minimization models in the route choice context. In *Proceedings of the 13th International Conference on Travel Behavior Research*, Toronto, Canada, 2012.