11-2016

A decomposition method for estimating recursive logit based route choice models

Tien MAI

Singapore Management University, atmai@smu.edu.sg

Fabian BASTIN

Emma FREJINGER

A Decomposition Method for Estimating Recursive Logit Based Route Choice Models

Tien Mai*, **Emma Frejinger, Fabian Bastin**

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. Fosgerau et al. (2013) recently proposed the recursive logit (RL) model for route choice problems that can be consistently estimated and easily used for prediction without any sampling of choice sets. Its estimation however requires solving many large-scale systems of linear equations, which can be computationally costly for real data sets. We design a decomposition (DeC) method in order to reduce the number of linear systems to be solved, opening the possibility to estimate more complex RL based models, such as, mixed RL models. We illustrate the approach on two mixed RL specifications, one using random coefficients and one incorporating error components associated with subnetwork (Frejinger and Bierlaire, 2007). The models are estimated on a real network with more than 3000 nodes and 7000 links, and a cross-validation study is performed. The results suggest that the DeC method significantly speeds up the estimation of the RL model and allows to estimate the mixed RL models in a reasonable time. The mixed RL model yields sensible parameter estimates and the fit and prediction are significantly better than the RL model.

Keywords: Decomposition method, route choice, mixed recursive logit models, subnetworks, cross-validation.

Acknowledgement: This research was funded by the National Sciences and Engineering Research Council of Canada (NSERC), discovery grant 435678-2013. We have benefited from valuable discussions with Mogens Fosgerau.

---

* Corresponding author: AnhTien.Mai@cirrelt.ca

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

1 Introduction

Given a transport network with a number of attributes associated with each link or path, the objective of route choice models is to assign a choice probability to each path in the network. Discrete choice models are typically used, despite two main issues, namely (i) choice sets of paths are unknown to the analyst and the set of all feasible paths for a given origin-destination pair cannot be enumerated in this context, and (ii) path utilities may be correlated, for instance, due to physical overlap in the network. This paper concerns recursive route choice models that can be consistently estimated without sampling of choice sets and allows links and paths to be correlated. The estimation of these models is expensive since it requires solving many systems of linear equations to compute the log-likelihood function and its gradients. We deal with this challenge by proposing a new method that allows to reduce the number of systems to be solved, saving computational time significantly.

Most of the existing route choice models are based on choice sets of paths that are sampled from the full choice sets. The sampled choice sets can be considered as the actual choice sets (e.g. Ben-Akiva and Bierlaire, 1999), or can be used to get consistent estimates by correcting the path choice probabilities (Frejinger et al., 2009, Lai and Bierlaire, 2015). Recently, Fosgerau et al. (2013) propose the recursive logit (RL) model, which considers the set of all feasible paths in the network. The choice probability for each path is computed by means of dynamic programming, the choice of a path being described as a sequence of link decisions. The model can be consistently estimated and is quickly used for prediction without sampling of choice sets. It however retains the well-know independence of irrelevant alternatives (IIA) property which is undesirable in a route choice setting (Mai et al., 2015b). Mai et al. (2015a) propose the nested RL (NRL) model that relaxes the IIA property over paths by assuming that scale parameters are link specific.

The estimation of the RL model requires solving a dynamic programming (DP) problem, which is considerably more time consuming than estimating the MNL model with finite choice sets. The DP problems can become costly to solve, e.g., for very large networks, large number of observations (destinations), or for models that require simulation (mixed logit). In order to address this issue, we propose a decomposition (DeC) method that allows to reduce the number of linear systems to be solved when tackling the DP problem of the RL model. This method speeds up the estimation and opens the possibility to estimate mixed RL models or RL models with nonlinear-in-parameters link utilities.

We apply the DeC to mixed recursive logit models. The mixed logit is attractive model since it relaxes the IIA assumption and is fully flexible, in the sense that it can approximate any random utility model (McFadden and Train, 2000). However, it is rarely used in the context of route choice analysis due its estimation cost. For example, Bekhor et al. (2002) estimate a mixed MNL model based on the Error Component (EC) approach (Bolduc and Ben-Akiva, 1991) using route choice data collected in Boston. Frejinger and Bierlaire (2007) also use the EC approach to model path correlations using subnetwork components. These approaches are however based on generated choice sets of paths. The mixed RL models considered in this paper take advantage of the RL model, as they can be consistently estimated and used for prediction without sampling choice sets, while allowing path and links to be correlated. The main concern is the computational cost of its estimation due to the presence of numerous linear systems. We show how the DeC method can be used for estimating mixed RL models and present estimation results for two mixed RL models based on the Borlänge network in Sweden.

The paper is structured as follows. Section 2 reviews the RL model and Section 3 introduces the DeC method. The mixed RL model is presented in Section 4. The subnetwork error components model is presented in Section 5. We provide in Section 6 numerical results based on real data, and finally, Section 7 concludes.

2 Recursive logit

The RL model proposed in Fosgerau et al. (2013) is based on the observation that a path choice can be formulated as a sequence of link choices and modeled in a dynamic discrete choice framework. A directed connected network (not assumed acyclic) $G = (\mathcal{A}; \mathcal{V})$ is considered, where $\mathcal{A}$ is the set of links and $\mathcal{V}$ if the set of nodes. For each link $k \in \mathcal{A}$, we denote the set of outgoing links from the sink node of k by $A(k)$. We assume that D destinations are present, and we associate an absorbing state with each destination by extending the network with a set of dummy links $\mathcal{D} = \{d_1, \ldots, d_D\}$, each d_l departing from the sink node of the destination l. The set of all links is denoted as $\tilde{\mathcal{A}} = \mathcal{A} \cup \mathcal{D}$. Figure 1 shows a simple network with 4 destinations i.e. d, c, e and f. Four dummy links d_1, d_2, d_3, and d_4, without successors, are added to the destinations, respectively.

Given a destination $d \in \mathcal{D}$ and two links $k \in \mathcal{A}$, $a \in \mathcal{A} \cup \{d\}$, $a \in A(k)$, the following instantaneous utility

$$u^n(a|k;\beta) = v^n(a|k;\beta) + \mu(\epsilon(a) - \gamma) \tag{1}$$

is associated with action $a \in A(k)$ of individual n, where β is a vector of parameters to be estimated, $v^n(a|k;\beta)$ is a deterministic utility, and γ is the Euler's constant. The

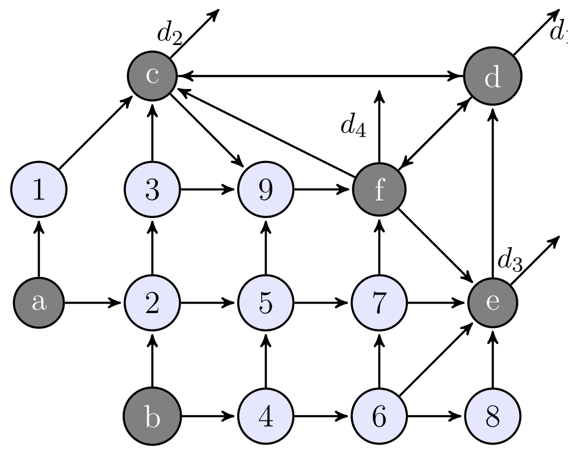

Figure 1: A small network with multiple destinations

deterministic utilities associated with destination d are set to zero, i.e., $v^n(d|k;\beta) = 0$. The random terms $\epsilon(a)$ are assumed to be i.i.d extreme value type I and the Euler's constant is used in order to ensure that the error terms have zero mean. We note that, in the NRL model (Mai et al., 2015a), the scales μ are assumed to be link specific so the model allows path utilities to be correlated in a fashion similar to the nested logit model (Ben-Akiva, 1973, McFadden, 1978). For notational simplicity, we omit from now an index for individual n but note that the utilities can be individual specific.

As discussed in Fosgerau et al. (2013), given a link $k \in \mathcal{A}$, the expected maximum utility $V^d(k;\beta)$, from link k until the destination d, is given recursively by the logsum

$$\frac{1}{\mu}V^d(k;\beta) = \ln\left(\sum_{a\in A(k)} e^{\frac{1}{\mu}(v(a|k;\beta)+V^d(a;\beta))}\right) \quad \forall k \in \mathcal{A},$$

and $V^d(d;\beta) = 0$ by assumption. The superscript d indicates that the value functions are destination specific. The choice probability of a given path $\sigma = \{k_0,\ldots,k_J\}$ is

$$P(\sigma;\beta) = e^{-V^d(k_0;\beta)} \prod_{i=1}^{J} e^{v(k_{i+1}|k_i;\beta)}. \tag{2}$$

The vector of parameters β can be estimated by maximizing the log-likelihood (LL) function defined over path observations. The method introduced by (Fosgerau et al.,

2013) to estimate the RL model requires to solve one linear system per destination, at each iteration of the LL maximization, but this is computationally too costly when dealing with complex models such as the mixed logit.

3 Decomposition method

In this section we propose a Decomposition (DeC) method that allows to consider only one system of linear equations when computing the value functions associated with all the destinations, thus significantly reducing the computational time of the estimation process.

Given a destination d, we introduce a destination-specific matrix M^d of size $(|\mathcal{A}|+1) \times (|\mathcal{A}|+1)$, whose elements are defined for all $k, a \in \mathcal{A} \cup \{d\}$ as

$$m_{ka} = \begin{cases} \delta(a|k)e^{\frac{1}{\mu}v(a|k)} & \text{if } k \in \mathcal{A} \\ 0 & \text{if } k = d, \end{cases}$$

where $\delta(a|k) = 1$ if $a \in A(k)$, and 0 otherwise. We impose that the last column and last row correspond to the dummy link d, so

$$M^d = \begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,|\mathcal{A}|+1} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,|\mathcal{A}|+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

A vector z^d of size $|\mathcal{A}|+1$ is also defined with elements $z^d_k = e^{\frac{1}{\mu}V(k)}$ for all states $k \in \mathcal{A} \cup \{d\}$ and b^d is a vector of size $|\mathcal{A}|+1$ with zero values for all states except for dummy d, the corresponding component being equal to one. Fosgerau et al. (2013) show that components of z are solutions to the following linear system

$$z^d = M^d z^d + b^d. \tag{3}$$

We now describe the DeC method based on the set of destinations taken from the observations. We first define a matrix M^0 of size $(|\mathcal{A}|+1) \times (|\mathcal{A}|+1)$ with entries

$$m^0_{ka} = \begin{cases} m_{ka} & \text{if } k, a \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}.$$

Moreover, we define a matrix U^d of size $(|\mathcal{A}|+1)\times(|\mathcal{A}|+1)$, for each absorbing state $d \in \mathcal{D}$, with entries

$$u^d_{ka} = \begin{cases} \delta(d|k)e^{\frac{1}{\mu}v(d|k)} & \text{if } a = d \\ 0 & \text{otherwise.} \end{cases}$$

So for each $d \in \mathcal{D}$ we have $M^d = M^0 + U^d$, or more explicitly

$$\underbrace{\begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,|\mathcal{A}|+1} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,|\mathcal{A}|+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{M^d} = \underbrace{\begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & 0 \\ m_{2,1} & m_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{M^0} + \underbrace{\begin{pmatrix} 0 & 0 & \cdots & m_{1,|\mathcal{A}|+1} \\ 0 & 0 & \cdots & m_{2,|\mathcal{A}|+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{U^d}.$$

Since $V^d(d) = 0$, $z^d_{|\mathcal{A}|+1} = e^{V^d(d)/\mu} = 1$ and

$$U^d z^d = \begin{pmatrix} 0 & 0 & \cdots & m_{1,|\mathcal{A}|+1} \\ 0 & 0 & \cdots & m_{2,|\mathcal{A}|+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} z^d_1 \\ z^d_2 \\ \vdots \\ z^d_{|\mathcal{A}|+1} \end{pmatrix} = \begin{pmatrix} m_{1,|\mathcal{A}|+1} \\ m_{2,|\mathcal{A}|+1} \\ \vdots \\ 0 \end{pmatrix},$$

we can therefore write (3) as

$$z^d = M^0 z^d + U^d z^d + b^d = M^0 z^d + t^d + b^d,$$

where t^d is the last column of the matrix U^d. Let Z be a matrix of size $(|\mathcal{A}|+1)\times|\mathcal{D}|$ whose columns are the vectors z^d, for all $d \in \mathcal{D}$

$$Z = [z^{d_1}, ..., z^{d_{|\mathcal{D}|}}],$$

and B be a matrix of size $(|\mathcal{A}|+1)\times|\mathcal{D}|$ whose columns are the vectors $t^d + b^d$, $\forall d \in \mathcal{D}$. We obtain a new system of linear equations

$$Z = M^0 Z + B$$

equivalent to

$$(I - M^0)Z = B. \tag{4}$$

The system has a solution if $I - M^0$ is invertible. The conditions to ensure that $I - M^d$ is invertible are discussed in Fosgerau et al. (2013), and they can be extended to $I - M^0$. In essence, the existence of a solution to the above system depends on the size of the scaled instantaneous utilities and on the balance between the number of paths connecting the nodes in the network. It is easy to find a feasible solution by using large enough magnitude of the model parameters.

For a given dummy $d_t \in \mathcal{D}$ and a link $k \in A$, the corresponding value function is

$$V^{d_t}(k) = \mu \ln(z_{kd_t}),$$

where z_{kd_t} is the element of matrix Z corresponding to column d_t and row k. Hence, the associated LL function is

$$LL(\beta) = \frac{1}{N} \sum_{n=1}^{N} \left(v(\sigma_n, \beta) - \ln z_{o_n d_n}(\beta) \right),$$

where $v(\sigma_n, \beta) = \sum_{i=0}^{I_n - 1} v(k^n_{i+1} | k^n_i)$ is the sum of the deterministic link utilities of path σ_n and o_n, d_n are the origin and dummy link of path σ_n, respectively. The gradient and Hessian of the LL function require the derivatives of the value functions. These can be obtained by deriving the Jacobian of vector Z with respect to a parameter β_i as

$$\frac{\partial Z}{\partial \beta_i} = (I - M^0)^{-1} \frac{\partial M^0}{\partial \beta_i} Z + (I - M^0)^{-1} \frac{\partial B}{\partial \beta_i},$$

and the Hessian with respect to two parameters β_i, β_j is

$$\frac{\partial^2 Z}{\partial \beta_i \partial \beta_j} = (I - M^0)^{-1} \left(\frac{\partial^2 M^0}{\partial \beta_i \partial \beta_j} Z + \frac{\partial M^0}{\partial \beta_i} \frac{\partial Z}{\partial \beta_j} + \frac{\partial M^0}{\partial \beta_j} \frac{\partial Z}{\partial \beta_i} + \frac{\partial^2 B}{\partial \beta_i \partial \beta_j} \right).$$

Hence, we only need to solve the system of linear equations (4) once, and vector Z contains all the value functions for all destinations.

Finally, it is important to note that the RL model and its extensions may be useful for traffic simulation. In this context, one needs to compute the next-link choice probabilities $P(a|k)$, $\forall k \in \mathcal{A}, a \in A(k)$, and link flows in the network. The next-link choice probabilities given by matrix P are computed using the value functions, and the link flows can be obtained by inverting matrix $I - P^T$, where I is the identity matrix. Baillon and Cominetti (2008) show that $I - P^T$ is invertible. The DeC method allows to compute matrix P quickly, hence it is useful for prediction and simulation, in addition to the maximum likelihood estimation.

4 Mixed recursive logit

The choice probability of a path given by the RL model can be computed using (2). In the mixed multinomial logit framework, the values $v(k_{i+1}|k_i; \beta)$ themselves contain random elements. Following Train (2003), we assume that β is derived from a random vector ω and a parameters vector θ, i.e. $\beta = \beta(\theta, \omega)$. For example, if β is a K-dimensional normally distributed random vector, whose components are mutually independent, we may chose $\omega = (\omega_1, \ldots, \omega_K)$ with $\omega_i \sim N(0, 1)$, $i = 1, \ldots, K$ and let θ specify the means and the standard derivations of the components of β. The unconditional choice probability is obtained by taking the expectation over the random coefficients

$$P(\sigma; \theta) = \mathbb{E}_\omega[P(\sigma; \beta)] = \int P(\sigma; \theta, \omega) f(\omega) d\omega, \tag{5}$$

where the expectation is taken with respect to ω and $f(\cdot)$ is its density function. The vector θ can be estimated by maximizing the LL function defined over the set of path observations $n = 1, \ldots, N$

$$LL(\theta) = \frac{1}{N} \sum_{n=1}^{N} \ln P(\sigma_n; \theta).$$

This involves the computation of $P(\sigma_n; \theta)$ for each observation, and therefore, by (5), one multidimensional integral per individual. An analytical expression of (**??**) can usually not be obtained, it therefore has to be numerically approximated, either by quadrature methods, or by simulation.

Assuming that the integral dimension is K, (5) is approximated as

$$\tilde{P}(\sigma_n; \theta) = \sum_{i=1}^{R_n} w_{n,i} P(\sigma_n; \theta, \omega_{n,i}), \tag{6}$$

where $\omega_{n,i}$ are the integration nodes and $w_{n,i}$ are the integrations weights. The nodes and weights can be deterministically produced or, for Monte-Carlo methods, randomly drawn from the distribution of ω, and (6) becomes

$$\tilde{P}(\sigma_n; \theta) = \frac{1}{R_n} \sum_{r_i=1}^{R_n} P(\sigma_n; \theta, \omega_{n,i}). \tag{7}$$

The Monte Carlo approach better scales with the integral dimension, and under mild conditions, the estimators derived from the simulated log-likelihood converge almost surely towards the true maximum likelihood estimators as the number of draws R^n tend to infinity (Bastin et al., 2006). Randomized quasi-Monte Carlo methods have also been considered for mixed logit models, while the improvement is sometimes limited (Munger et al., 2012).

A major burden is however that different integration nodes produce different linear systems of the form (4), that have to be solved, inducing a significant numerical cost increase. We can limit it by requiring the integration nodes to be the same among the observations, i.e.

$$w_{n,i} \equiv w_i, \quad \omega_{n,i} \equiv \omega_i, \quad n = 1, \dots, N.$$

The integration error still converges to zero as the number of integration is growing to infinity, the convergence being almost sure in the case of Monte Carlo methods. In the latter case, the simulated choice probabilities are no longer independent between the observations, resulting in an additional simulation bias that nevertheless goes to zero asymptotically with the number of draws, and in practice, it can often be neglected.

In order to illustrate this point, we compare the estimates obtained by the two approaches using the small network in Figure 1. The network is composed of 15 nodes and 28 arcs. There are two origin nodes a, b and four destination nodes which are c, d, e, f. Note that each origin connects to all destinations, so there are eight OD pairs in total. Travel time is the only attribute considered in our example and they are generated uniformly in interval $[0, 1]$. We note that dummy links d_1, d_2, d_3, d_4 are added to destinations c, d, e, f, respectively for the estimation of the mixed RL model. Following the simulation literature, we say that we use independent random numbers (IRN) when the values $\omega_{n,i}$ are independently drawn between the route choice observations, and common random numbers (CRN) when we use the same set of random draws between the observations.

We assume that each link a is associated with a travel time $TT(a)$ and we consider the following instantaneous utility with respect to link a given link k

$$u(a|k) = \beta_1 TT(a) + \beta_2 + (\epsilon(a) - \gamma), \tag{8}$$

where $\epsilon(a)$ follows an extreme value type I, γ is the Euler's constant, β_1 and β_2 are normally distributed. We assume that $\beta_1 = N(\theta_{TT}, \sigma^2_{TT})$ and $\beta_2 = N(0, \sigma^2)$. The path observations are simulated using parameters $\{\theta^0_{TT}, \sigma^0_{TT}, \sigma^0\} = \{-11, 0.1, 0.1\}$. 5 paths are simulated for each OD pair, so the sample contains 40 path observations in total. We note that the number of observations is kept small because we do not aim at recovering the true parameters and even with small number of observations, the LL function is well defined.

We estimate the mixed RL model with CRN and IRN approaches. All the reported estimated parameters and final LL values are based on 100 independent simulations. For each estimation with CRN or IRN we use 50, 200 and 1000 Monte

Carlo draws. The means and standard derivations over 100 simulations are reported in Table 1. We also compute a t-test for the null hypothesis that the mean of the estimated parameters given by the CRN is not significant different from its respective value given by the IRN approach. The results show that the estimated parameters as well as the final LL given by the CRN method are not significantly different with the ones given by the IRN method at the 0.05 significance level. As expected, the standard deviations are larger for the lower numbers of draws. The parameter estimates of CRN and IRN can be different (e.g. 0.089 versus 0.461) without being significantly different. When the number of draws is large enough (e.g. 1000) the LL and parameter estimates have similar values and small standard deviations. In summary, for this example, the CRN method is an alternative to the standard IRN.

	Number of draws	50		200		1000	
		CRN	IRN	CRN	IRN	CRN	IRN
$LL(\hat{\beta})$	mean	0.699	0.698	0.700	0.699	0.701	0.701
	std. dev.	0.002	0.003	0.001	0.001	0.0005	0.0008
	t-test	0.137	-	0.433	-	0.137	-
$\hat{\theta}_{TT}$	mean	-12.433	-11.810	-11.787	-11.667	-11.501	-11.410
	std. dev.	0.554	0.597	0.289	0.440	0.136	0.400
	t-test	-1.044	-	-0.273	-	-0.226	-
$\hat{\sigma}_{TT}$	mean	0.284	0.971	0.089	0.461	0.107	0.193
	std. dev.	0.161	0.683	0.077	0.437	0.069	0.184
	t-test	-1.007	-	-0.849	-	-0.465	-
$\hat{\sigma}$	mean	0.670	0.479	0.553	0.487	0.501	0.417
	std. dev.	0.121	0.110	0.069	0.113	0.035	0.165
	t-test	1.730	-	0.585	-	0.514	-

Table 1: Estimation results with IRN and CRN

Finally, we note that, thanks to the DeC method, the computational time for estimating the model using the CRN is from 10 to 50 times less than when estimating the model using the IRN. We also observe that, for the CRN, the optimization algorithm often converges faster in terms of number of iterations, compared to the conventional IRN.

5 Modeling correlations with subnetwork components

In this section we present a route choice modeling approach that is convenient to use with the mixed RL model. It was proposed by Frejinger and Bierlaire (2007), and it as called the subnetwork approach. A subnetwork is a set of components where each component is defined as a sequence of links. A component can, for instance, corresponds to a main road in the transport network. The approach is based on the assumption that paths may be correlated if they share a subnetwork component. To capture this correlation, Frejinger and Bierlaire (2007) propose an Error Component (EC) model (see for instance Bekhor et al., 2002). Inspired by this approach, in the following, we propose an EC model for the mixed RL model that allows to capture the correlation between paths and links that share subnetwork components in the network.

Assuming that the network is composed of Q subnetwork components. We add error component factors to the instantaneous utilities as

$$u(a|k;\beta,\sigma) = v(a|k;\beta) + F(a)^{\mathrm{T}}T(\sigma)\zeta + \mu(\epsilon(a) - \gamma),\ \forall k \in \mathcal{A}, a \in A(k),$$

where $F(a)$ is a vector of dimension Q and each element $F(a)_q$ associates link a in overlaps with subnetwork component q (i.e. if link a is a part subnetwork component q, $F(a)_q = \sqrt{l(a)}$ and $F(a)_q = 0$ otherwise, where $l(a)$ is the length of link a), $T(\sigma) = \mathrm{diag}(\sigma_1, \ldots, \sigma_Q)$ where σ is a vector of covariance parameters to be estimated, and ζ is a vector of size Q where each element is a $N(0,1)$ random variable.

6 Numerical results

In this section we use the same data and network used in Frejinger and Bierlaire (2007), Fosgerau et al. (2013) and Mai et al. (2015a) to apply the DeC method and provide estimation and prediction results for the mixed RL models with random parameters and error components. The network consists of of 3077 nodes and 7459 links, and with static and deterministic travel times. The sample consists of 1832 trips corresponding to simple paths with a minimum of five links. There are 466 destinations, 1420 different origin-destination (OD) pairs and more than 37,000 link choices in this sample. For the sake of comparison, we also estimate the RL models (Fosgerau et al., 2013).

As in Frejinger and Bierlaire (2007), the subnetwork components of the Borlänge network are defined based on the main roads for traversing the city center. Two of the Swedish national roads traverse Borlänge. The subnetwork is composed of these national roads (referred to as R.50 and R.70) and there are two subnetwork components for each national road (north and south directions), leading to a total of four subnetwork components, denoted R.70N, R.70S, R.50S and R.50N. In addition, there is one component for the road segment in the city center which is denoted RC.

Figure 2 shows the Borlänge network and 5 subnetwork components. Reader can consult Frejinger and Bierlaire (2007) for more details.

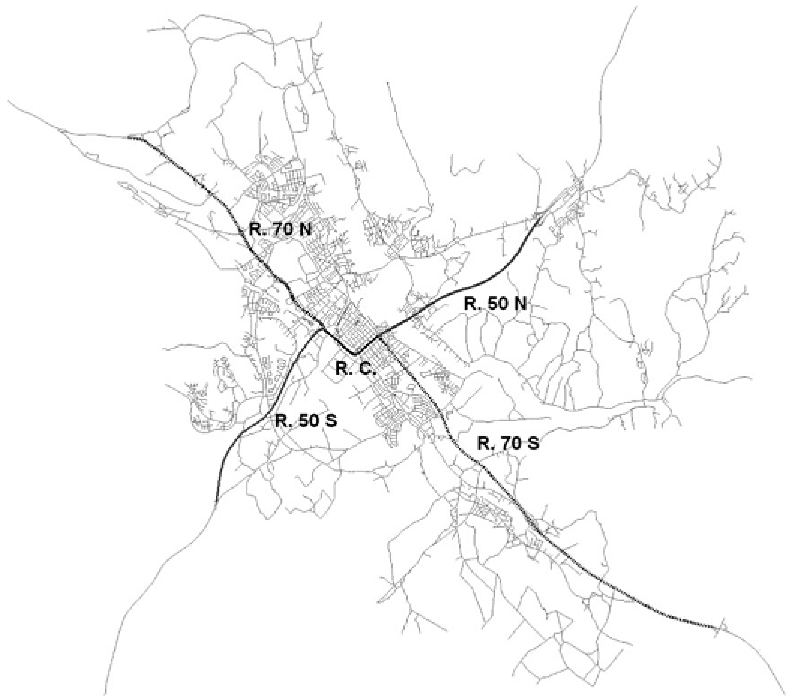

Figure 2: Borlänge subnetworks (Frejinger and Bierlaire, 2007)

6.1 Specification of models

The same four attributes as in Fosgerau et al. (2013) are used in the instantaneous utilities: link travel time $TT(a)$, a left turn dummy $LT(a|k)$ that equals to one if the turn angle from link k to a is larger than 40 degrees and less than 177 degrees, a u-turn dummy $UT(a|k)$ that equals to one if the turn angle is larger than 177, and a link constant $LC(a)$ set to 1 for all links in our experiments. Fosgerau et al. (2013) also propose the link size (LS) attribute for overlapping paths. This attribute is however origin-destination specific and is not compatible with the DeC model, therefore we do not use this attribute for the mixed RL models.

We specify the instantaneous utilities for different models for link a given link k, $a \in A(k)$ as

- RL

$$u^{RL}(a|k;\beta) = \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) + (\epsilon(a) - \gamma),$$

- RL with the LS attribute (RL-LS)

$$u^{RL-LS}(a|k;\beta) = \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) + \beta_{LS}LS(a) + (\epsilon(a) - \gamma),$$

- mixed RL with random parameters (MRL-RP)

$$u^{MRL}_{RP}(a|k;\beta) = \beta^*_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) + (\epsilon(a) - \gamma),$$

- mixed RL with EC (MRL-EC)

$$u^{MRL}_{EC}(a|k;\beta,\sigma) = \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) + F(a)^{\mathrm{T}}T(\sigma)\zeta + (\epsilon(a) - \gamma),$$

- mixed RL with EC and random parameters (MRL-REC)

$$u^{MRL}_{REC}(a|k;\beta,\sigma) = \beta^*_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) + F(a)^{\mathrm{T}}T(\sigma)\zeta + (\epsilon(a) - \gamma),$$

where $\epsilon(a)$ are i.i.d standard extreme value type I, β^*_{TT} is specified to be normal distribution $\beta^*_{TT} \sim N(\mu_{TT}, \sigma^2_{TT})$, ζ is a vector of size 5 where each element is a $N(0,1)$ random variable, and $F(a)$ is a vector of size 5 where each element $F(a)_q$ with respect to link a and subnetwork component q is defined as

$$F(a)_q = \begin{cases} 0 & \text{if } a \notin q \\ \sqrt{TT(a)} & \text{otherwise.} \end{cases}$$

The vector of the covariance parameters is $\sigma = \{\sigma_{R50N}, \sigma_{R50S}, \sigma_{R70N}, \sigma_{R70S}, \sigma_{RC}\}$.

6.2 Estimation results

For the estimation of the the mixed RL model we adopt the nested fixed point algorithm designed by Rust (1987). This algorithm combines an outer iterative non-linear optimization algorithm for searching over the parameter space with an inner algorithm for solving the value functions. Note that the value functions are solved for each MC draw using the DeC method and we use the basic trust region method (for instance Nocedal and Wright, 1999) with BFGS (Broyden, 1970, Fletcher, 1970, Goldfarb, 1970, Shanno, 1970) as the non-linear optimization algorithm.

We estimate the mixed RL models (MRL-RP, MRL-EC, MRL-REC) and report the results based on 10 independent simulations and for each simulation we use $R = 500$ Monte Carlo random draws. The parameter estimates are reported in the appendix, Tables 6, 7, and 8. The β estimates are comparable to those previously published using the same data. The parameter estimates have their expected signs and are all highly significantly different from zero. Over 10 simulations, the β estimates are closer in values, compared to the σ estimates. We note that Frejinger and Bierlaire (2007) estimate path-based models using the same data and the sub-network approach, and report that all the σ estimates are significantly different from zero except $\hat{\sigma}_{R50S}$.

We plot the final LL values over 10 repetitions in Figure 3. The averaged final LL values over 10 MC repetitions are reported in Table 2. For the sake of comparison we also include the RL models (RL and RL-LS). We only report the final LLs given by these models and reader can consult Fosgerau et al. (2013) for more details of the estimation results. The mixed RL models with EC performs better than the RL models in fit (the likelihood ratio test results are reported in Table 3 based on the averaged final LL values in Table 2).

Before presenting the prediction results, we make some remarks on computational time. The code is implemented in MATLAB 2013a (available upon request) and we use an Intel(R) machine, Core^{TM} i5-3210M CPU 2.50GHz. We use the trust region algorithm with BFGS to estimate all the models. The estimation of the mixed RL models is much more costly, compared to the RL models. We need from 3 to 5 days to estimate the mixed RL models, while it takes half day for estimating the RL models.

	RL	RL-LS	MRL-RP	MRL-EC	MRL-REC
$LL(\widehat{\beta})$	-6303.90	-6045.60	-6229.69	-5970.29	-5921.61

Table 2: Averaged final log-likelihood values

6.3 Prediction results

In this section we use a cross validation approach to compare the prediction performance of the different models considered above, i.e, the RL and mixed RL models.

Models	χ^2	p-value
RL & MRL-RP	148.42	3.84e-34
RL & MRL-EC	667.22	6.00e-142
MRL-RP & MRL-REC	616.16	6.52e-131
MRL-EC & MRL-REC	97.36	5.78e-23

Table 3: Likelihood ratio test results

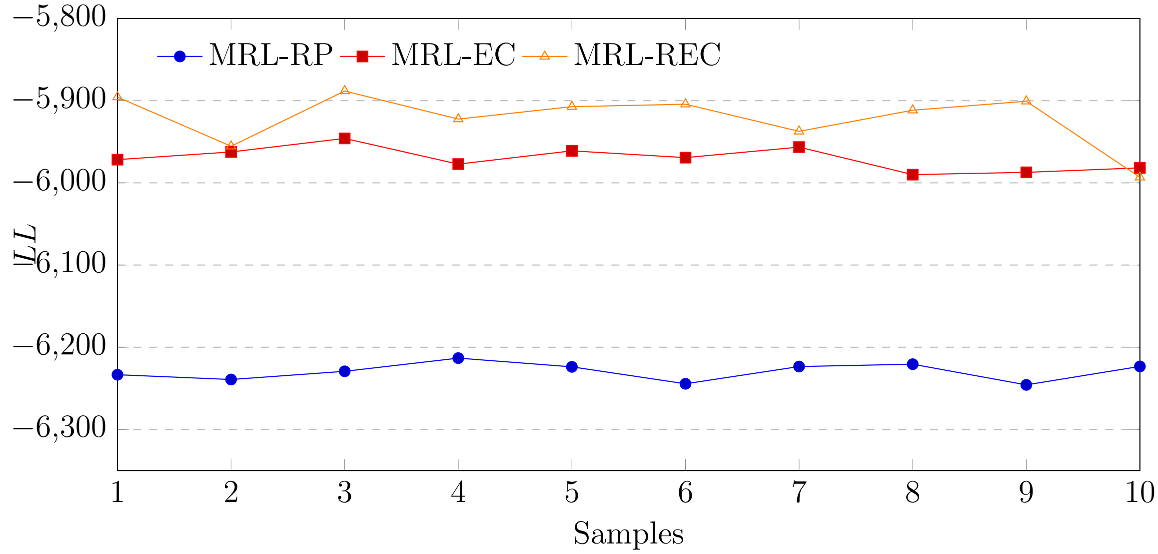

Figure 3: Final log-likelihood over 10 MC repetitions

The real sample (1832 observations) is divided into two sets by drawing observations uniformly: one set of 80% of observations is used for estimation, and one set of the remaining 20% observations is used as holdout to evaluate the predicted probabilities. We generate 20 different holdout samples of the same size by reshuffling the real sample. The LL loss values then are used to evaluate the prediction performance.

For each holdout sample i, $0 \leq i \leq 20$, we estimate the parameters $\hat{\theta}_i$ based on the respective estimation sample, and we compute the test errors err_i using the holdout sample and these parameter estimates as

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

where PS_i is the set of observations corresponding to holdout sample i, and $|PS_i|$ is the size of PS_i. Indeed, err_i depends on PS_i. Similarly to Mai et al. (2015a), we compute the average of err_i values over samples in order to have unconditional test error values

$$\overline{err}_p = \frac{1}{p}\sum_{i=1}^{p} err_i \quad \forall 1 \leq p \leq 20. \tag{9}$$

We plot the values of $\overline{err}_p$, $1 \leq p \leq 20$ in Figure 4 (lower the value, the better is the performance) and in Table 4, we report the average of the test error values given by five models over 20 samples. Indeed, the value of $\overline{err}_p$ becomes stable when p increases for each model. The mixed RL models are better than the RL model in term of prediction. The MRL-REC has a significant better fit and also a better prediction performance, compared to the other models.

RL	RL-LS	MRL-RP	MRL-EC	MRL-REC
3.39	3.34	3.36	3.22	3.19

Table 4: Average of test error values over 40 holdout samples

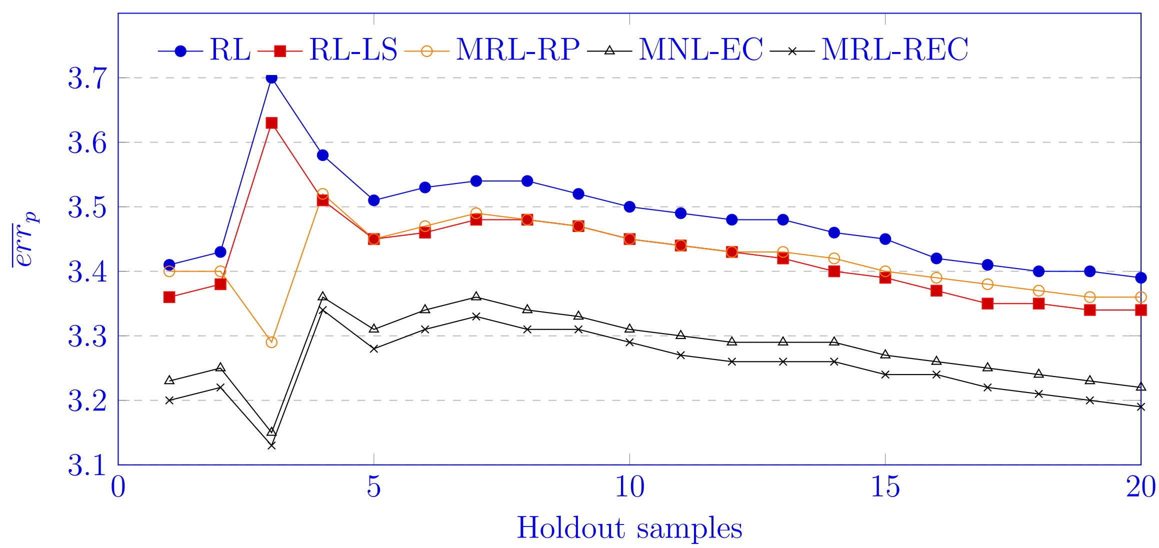

Figure 4: Average of the test error values over holdout samples

6.4 Evaluation of computational time for the decomposition method

In this section, we provide a comparison of computational time to show how the DeC method speeds up the estimation of the RL model. We compare the performance of the DeC method with the original approach based on the same data used in the previous section.

For the sake of comparison we estimate the path-based logit models proposed by Frejinger et al. (2009) based on the same data set. The models require sampling of choice sets and we draw 50 samples for each OD pair observation. The computational times associated with evaluating the LL and gradients are reported for two models, one with and one without the Extended Path Size attribute (Frejinger et al., 2009). We denote them PL and PSL, respectively. Recall that all the models are estimated with 1832 path observations (466 destinations).

The code is also implemented in MATLAB and it has not been parallelized. The computational times are reported in Table 5. For each "LL and gradients" evaluation, the DeC method is about 30 times faster, compared to the original method. The computations associated with the path-based logit models are also fast because the LL can be directly evaluated without computing the value functions. However, the total computational time for the estimation of the path-based models is high because choice sets need to be sampled before estimating. Finally, we note that the RL model with LS attribute is costly to estimate, compared to the model without.

	DeC method	Original method		Path-based models	
Models	RL	RL	RL-LS	PL	PSL
LL & gradients	10	284	1274	5.28	7.6
Estimation	230	18279	29638	3560	4146

Table 5: *Comparison of computational times (in seconds)*

7 Conclusion

In this paper we proposed a decomposition (DeC) method to speed up the estimation of the RL modelss. We have shown that the DeC approach can be used to estimate the mixed RL model in reasonable time. We provided estimation and prediction results using a real data set. We estimated the mixed RL models based on two different approaches which use random parameters and EC using subnetwork components. The results showed that the MRL-REC model has significantly better fit and better prediction than the RL models (with and without LS attribute) and other mixed RL models.

The DeC method significantly speeds up the estimation of the RL model. Our next steps will be dedicated to applying and testing the DeC method to other complex extensions of the RL model, e.g., the NRL model Mai et al. (2015a), the RL model for stochastic time-dependent networks and regret-based RL models. We are also interested in applying the mixed RL to other networks with panel data and different EC models.

Acknowledgements

This research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC), discovery grant 435678-2013. We have benefited from valuable discussions with Mogens Fosgerau.

References

Baillon, J.-B. and Cominetti, R. Markovian traffic equilibrium. *Mathematical Programming*, 111(1-2):33–56, 2008.

Bastin, F., Cirillo, C., and Toint, Ph. L. Convergence theory for nonconvex stochastic programming with an application to mixed logit. *Mathematical Programming, Series B*, 108(2–3):207–234, 2006.

Bekhor, S., Ben-Akiva, M., and Ramming, M. Adaptation of logit kernel to route choice situation. *Transportation Research Record*, 1805:78–85, 2002.

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.

Bolduc, D. and Ben-Akiva, M. A multinomial probit formulation for large choice sets. In *Proceedings of the 6th International Conference on Travel Behaviour*, volume 2, pages 243–258, 1991.

Broyden, C. G. The convergence of a class of double-rank minimization algorithms 1. General considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.

Fletcher, R. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 1970.

Fosgerau, M., Frejinger, E., and Karlström, A. A link based network route choice model with unrestricted choice set. *Transportation Research Part B*, 56(1):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.

Goldfarb, D. A family of variable metric updates derived by variational means. *Mathematics of Computation*, 24(109):23–26, 1970.

Lai, X. and Bierlaire, M. Specification of the cross nested logit model with sampling of alternatives for route choice models. *Transportation Research Part B*, 80:220–234, 2015.

Mai, T., Fosgerau, M., and Frejinger, E. A nested recursive logit model for route choice analysis. *Transportation Research Part B*, 75:100–112, 2015a.

Mai, T., Frejinger, E., and Bastin, F. A misspecification test for logit based route choice models. *Economics of Transportation*, 4(4):215–226, 2015b.

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.

McFadden, D. L. and Train, K. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–270, 2000.

Munger, D., L'Ecuyer, P., Bastin, F., Cirillo, C., and Tuffin, B. Estimation of the mixed logit likelihood function by randomized quasi-monte carlo. *Transportation Research Part B*, 46(2):305–320, 2012.

Nocedal, J. and Wright, S. J. *Numerical Optimization*. Springer, New York, NY, USA, 1999.

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

Shanno, D. F. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656, 1970.

Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2003.

Appendix

	1	2	3	4	5	6	7	8	9	10
$\widehat{\mu}_{TT}$	-2.845	-2.855	-2.881	-2.964	-3.035	-2.690	-2.967	-3.028	-2.675	-2.992
Rob. Std. Err.	0.106	0.107	0.108	0.113	0.114	0.101	0.112	0.114	0.100	0.113
Rob. t-test(0)	-26.74	-26.71	-26.69	-26.28	-26.64	-26.76	-26.58	-26.53	-26.70	-26.57
$\widehat{\sigma}_{TT}$	0.800	0.688	0.697	-1.058	0.811	0.606	0.817	0.858	0.845	-0.810
Rob. Std. Err.	0.032	0.027	0.028	0.047	0.033	0.024	0.034	0.036	0.022	0.033
Rob. t-test(0)	25.02	25.19	24.49	-22.42	24.35	25.31	24.33	24.05	38.41	-24.20
$\widehat{\beta}_{LT}$	-0.928	-0.928	-0.927	-0.926	-0.927	-0.929	-0.926	-0.926	-0.929	-0.927
Rob. Std. Err.	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
Rob. t-test(0)	-30.18	-30.30	-30.09	-29.79	-29.99	-30.39	-29.96	-29.91	-30.41	-30.01
$\widehat{\beta}_{LC}$	-0.425	-0.424	-0.426	-0.430	-0.427	-0.423	-0.427	-0.428	-0.423	-0.427
Rob. Std. Err.	0.012	0.012	0.013	0.013	0.013	0.012	0.013	0.013	0.012	0.013
Rob. t-test(0)	-34.02	-33.93	-33.91	-33.70	-33.93	-33.95	-33.89	-33.93	-33.93	-33.92
$\widehat{\beta}_{UT}$	-4.400	-4.405	-4.396	-4.382	-4.392	-4.408	-4.391	-4.389	-4.410	-4.391
Rob. Std. Err.	0.115	0.115	0.115	0.115	0.115	0.115	0.115	0.115	0.115	0.115
Rob. t-test(0)	-38.31	-38.39	-38.26	-38.04	-38.19	-38.44	-38.19	-38.15	-38.46	-38.18
Final LL	-6233.45	-6239.26	-6229.30	-6213.25	-6223.86	-6244.47	-6223.50	-6220.64	-6245.80	-6223.35

Table 6: *Estimation results for the mixed RL with random parameters*

	1	2	3	4	5	6	7	8	9	10
$\widehat{\beta}_{TT}$	-2.873	-2.888	-2.885	-2.884	-2.913	-2.875	-2.903	-2.874	-2.890	-2.882
Rob. Std. Err.	0.118	0.119	0.118	0.119	0.120	0.118	0.119	0.119	0.120	0.117
Rob. t-test(0)	-24.26	-24.36	-24.49	-24.19	-24.25	-24.39	-24.34	-24.21	-24.05	-24.54
$\widehat{\beta}_{LT}$	-0.947	-0.945	-0.946	-0.949	-0.942	-0.946	-0.947	-0.941	-0.949	-0.943
Rob. Std. Err.	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
Rob. t-test(0)	-30.45	-30.42	-30.50	-30.70	-30.36	-30.56	-30.46	-30.58	-30.57	-30.46
$\widehat{\beta}_{LC}$	-0.406	-0.406	-0.406	-0.403	-0.403	-0.405	-0.405	-0.404	-0.403	-0.405
Rob. Std. Err.	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
Rob. t-test(0)	-28.38	-28.72	-28.50	-28.32	-28.20	-28.53	-28.46	-28.36	-28.14	-28.71
$\widehat{\beta}_{UT}$	-4.388	-4.386	-4.383	-4.394	-4.386	-4.391	-4.384	-4.391	-4.389	-4.391
Rob. Std. Err.	0.116	0.117	0.116	0.116	0.116	0.116	0.116	0.116	0.116	0.117
Rob. t-test(0)	-37.74	-37.64	-37.68	-37.75	-37.74	-37.71	-37.73	-37.78	-37.76	-37.53
$\widehat{\sigma}_{R50N}$	1.700	1.893	1.630	1.736	2.055	1.829	1.756	1.712	2.009	1.601
Rob. Std. Err.	0.045	0.045	0.033	0.038	0.072	0.038	0.041	0.042	0.077	0.035
Rob. t-test(0)	37.62	42.41	49.20	45.31	28.53	48.10	42.54	41.05	26.09	45.11
$\widehat{\sigma}_{R50S}$	1.796	1.553	1.663	1.661	1.482	1.487	1.781	1.942	1.702	1.892
Rob. Std. Err.	0.155	0.079	0.036	0.074	0.145	0.054	0.189	0.408	0.088	0.258
Rob. t-test(0)	11.61	19.71	46.19	22.57	10.25	27.49	9.41	4.75	19.23	7.32
$\widehat{\sigma}_{R70N}$	1.789	1.636	1.472	1.671	1.420	1.514	1.718	1.645	1.816	1.446
Rob. Std. Err.	0.074	0.076	0.033	0.077	0.044	0.034	0.077	0.034	0.111	0.043
Rob. t-test(0)	24.19	21.58	45.19	21.64	32.27	45.16	22.24	48.28	16.36	33.59
$\widehat{\sigma}_{R70S}$	1.567	1.802	1.881	1.543	1.701	1.433	1.759	1.525	1.338	1.614
Rob. Std. Err.	0.035	0.040	0.042	0.035	0.038	0.032	0.040	0.034	0.030	0.036
Rob. t-test(0)	44.75	44.58	45.02	44.49	44.30	44.83	44.46	44.63	44.10	45.01
$\widehat{\sigma}_{RC}$	2.464	2.869	2.643	2.825	2.855	2.762	2.525	2.692	2.762	2.513
Rob. Std. Err.	0.051	0.161	0.237	0.081	0.138	0.149	0.045	0.144	0.099	0.037
Rob. t-test(0)	48.09	17.82	11.16	34.82	20.76	18.51	55.53	18.76	27.78	67.82
Final LL	-5971.71	-5962.38	-5945.92	-5977.19	-5961.04	-5969.28	-5956.51	-5989.93	-5987.23	-5981.71

Table 7: *Estimation results for the mixed RL model with EC*

	1	2	3	4	5	6	7	8	9	10
$\widehat{\mu}_{TT}$	-3.458	-3.124	-3.451	-3.315	-3.264	-3.467	-3.333	-3.382	-3.489	-2.896
Rob. Std. Err.	0.134	0.125	0.134	0.123	0.121	0.131	0.122	0.122	0.133	0.117
Rob. t-test(0)	-25.873	-25.089	-25.840	-26.904	-26.953	-26.484	-27.351	-27.741	-26.271	-24.665
$\widehat{\sigma}_{TT}$	0.886	0.604	0.910	0.755	0.657	1.106	1.052	0.932	0.875	0.233
Rob. Std. Err.	0.079	0.109	0.041	0.029	0.025	0.046	0.045	0.037	0.034	0.006
Rob. t-test(0)	11.161	5.525	22.328	25.878	26.605	24.114	23.175	24.908	25.618	42.045
$\widehat{\beta}_{LT}$	-0.941	-0.943	-0.933	-0.936	-0.945	-0.939	-0.937	-0.931	-0.947	-0.946
Rob. Std. Err.	0.031	0.031	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.031
Rob. t-test(0)	-30.144	-29.978	-29.351	-29.613	-29.417	-29.232	-29.375	-29.372	-29.799	-30.634
$\widehat{\beta}_{LC}$	-0.413	-0.406	-0.422	-0.411	-0.416	-0.417	-0.420	-0.417	-0.418	-0.405
Rob. Std. Err.	0.015	0.015	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
Rob. t-test(0)	-28.014	-26.730	-30.222	-29.057	-29.862	-30.188	-29.089	-29.930	-30.304	-28.745
$\widehat{\beta}_{UT}$	-4.320	-4.370	-4.301	-4.348	-4.328	-4.319	-4.331	-4.318	-4.310	-4.395
Rob. Std. Err.	0.115	0.117	0.116	0.116	0.116	0.116	0.116	0.116	0.116	0.116
Rob. t-test(0)	-37.474	-37.438	-36.924	-37.451	-37.318	-37.132	-37.188	-37.153	-37.075	-37.806
$\widehat{\sigma}_{R50N}$	1.787	2.084	1.744	1.933	1.655	2.100	1.976	2.221	1.771	1.838
Rob. Std. Err.	0.107	0.134	0.055	0.053	0.033	0.059	0.037	0.059	0.039	0.031
Rob. t-test(0)	16.673	15.543	31.949	36.286	50.902	35.573	53.667	37.798	44.850	59.29
$\widehat{\sigma}_{R50S}$	1.520	1.590	1.331	1.644	1.574	1.332	1.832	1.227	1.681	1.543
Rob. Std. Err.	0.228	0.122	0.046	0.267	0.215	0.093	0.228	0.056	0.253	0.090
Rob. t-test(0)	6.670	13.069	26.490	6.153	7.315	14.328	8.052	21.993	6.635	17.213
$\widehat{\sigma}_{R70N}$	1.629	1.898	1.579	1.755	1.721	1.624	1.644	1.697	1.528	1.545
Rob. Std. Err.	0.030	0.096	0.049	0.020	0.101	0.031	0.056	0.043	0.036	0.041
Rob. t-test(0)	54.30	19.778	31.998	87.75	17.084	52.399	29.521	39.173	42.637	37.261
$\widehat{\sigma}_{R70S}$	1.777	1.326	1.693	1.710	1.725	1.356	1.489	1.501	1.696	1.379
Rob. Std. Err.	0.059	0.100	0.043	0.033	0.037	0.029	0.030	0.030	0.039	0.032
Rob. t-test(0)	29.934	13.193	38.987	51.974	46.881	46.479	49.535	49.761	43.927	43.763
$\widehat{\sigma}_{RC}$	2.833	2.928	2.763	3.080	2.843	2.977	2.645	2.702	3.012	2.805
Rob. Std. Err.	0.037	0.094	0.124	0.269	0.092	0.178	0.039	0.062	0.231	0.149
Rob. t-test(0)	76.56	31.149	22.285	11.447	30.990	16.770	62.011	38.562	13.039	18.874
Final LL	-5895.62	-5955.43	-5888.26	-5922.38	-5907.29	-5904.32	-5937.29	-5911.62	-5900.71	-5993.18

Table 8: *Estimation results for the mixed RL model with EC and RP*