Route choice analysis considering driving comfort and travel time: a random regret minimization model with heterogeneity

Weiliang Zeng*
Postdoctoral Research Fellow
Institution of Materials and Systems for Sustainability, Nagoya University
Furo-cho, Chikusa, Nagoya, Japan 464-8603
Tel: +81-52-789-3730
Fax: +81-52-789-5728
E-mail: zengweiliang@imass.nagoya-u.ac.jp
* Corresponding author

Tomio Miwa
Associate Professor
EcoTopia Science Institute, Nagoya University
Furo-cho, Chikusa, Nagoya, Japan 464-8603
Tel: +81-52-789-5818
Fax: +81-52-789-5728
E-mail: miwa@nagoya-u.jp

Takayuki Morikawa
Professor
Institute of Innovation for Future Society, Nagoya University
Furo-cho, Chikusa, Nagoya, Japan 464-8603
Tel: +81-52-789-3564
Fax: +81-52-789-5728
E-mail: morikawa@nagoya-u.jp
ABSTRACT

Route choice model has shown an essential role in many transport applications such as transport planning and network simulation. Route choice behavior strongly depends on the choice set, route characteristics and individual heterogeneity. With the development of regret theory, the random regret minimization (RRM) has gathered a lot of interest in route choice modeling. In this study, we aim to explore the route choice behavior from GPS and OBD (On-Board Diagnostics) data. The observed heterogeneity is explicitly incorporated into the Path-size random regret minimization model. In contrast with previous study, the driving comfort represented by heart rate is first considered as one of route attributes. Trip records include GPS trajectory and driving operation (OBD record) collected by private vehicles in Toyota city, Japan, are used to estimate the choice models. The estimation result indicates that driving comfort has a stronger effect than travel time on the route choice behavior. It shows that the random regret minimization based models perform better than the random utility maximization based models in this empirical analysis. It also confirms that incorporation of observed heterogeneities such as OD distance, age, gender, departure time and day of week have significant effects on the taste to travel time and driving comfort.

Keywords: Route choice behaviour, choice set generation, driving comfort, travel time, GPS data
1. Introduction

Route choice modeling is one of the crucial issues in urban transportation system, which plays a core role in both traffic assignment and network simulation. The route choice is a reflection of potential preferences for each available route and we usually assume that traveler chooses the best route by maximizing their utility. There are often multiple routes between an origin-destination (OD) pair. The utility of each route depends on route attributes and individual heterogeneity (Papinski et al., 2009; Li et al., 2016; Ciscal-Terry et al., 2016). Generally, the route choice probability can be estimated based on random utility maximization (RUM) theory (Ben-Akiva and Lerman, 1985). And many studies have explored the effect of route attributes such as travel time, distance and cost on the route choice behavior by using discrete choice model based on RUM theory (Prato, 2009). With the development of regret theory, the interest in random regret minimization (RRM) as an alternative paradigm to RUM has grown in recent years (Chorus, 2010; Chorus, 2012). The discrete choice paradigm of RRM has been applied to various choice contexts such as travel mode, parking route choice, automobile fuel choice, vehicle type, shopping destination, etc. (Chorus, 2010; Chorus, 2012; Hensher et al., 2013). In contrast with RUM, the criterion of RRM aims to obtain a solution minimizing a user’s regret. It is plausible to describe the route choice behavior with the assumption that drivers are regret minimizer who might prefer the route with minimum regret compared with others, rather than choose the route with maximum utility (Prato, 2014).

However, only limited studies have applied RRM to route choice analysis and comparison between RRM and RUM (Chorus et al., 2013). Furthermore, few studies based on RRM theory incorporated the individual heterogeneity due to the difficulty in data collection, though it is widely accepted that individuals have varying tastes for specific attributes. In this study, we aim to model travelers’ route choice behavior in the context of average travel time and driving comfort based on RRM. There are two contributions in contrast with previous research. First, we incorporate the individual heterogeneity to the Path-size RRM model, in which the OD pair specific heterogeneity, personal specific heterogeneity, and time specific heterogeneity are considered. Second, the driving comfort represented by heart rate is explicitly incorporated into the route attributes. Previous studies usually regarded driving comfort as an unobserved attribute since it was not easy to capture, even though it should be regarded as an important factor to route choice. We try to fill this gap by modeling the driving comfort that related to driving environment and driving operation in the data collection process.

The rest of the paper is structured as follows. Section 2 reviews the related work in route choice modeling. Section 3 gives the specification of the RRM based model. Section 4 describes the data collection for route choice modeling. Section 5 shows the estimation results. Last, we conclude with some perspectives for future research.

2. Literature review

Route choice has been widely investigated from both methodology and practice. The discrete choice paradigm of RUM is the most popular approach for route choice modeling because of the simple specification which makes the models easy to develop, estimate, and apply (Ben-Akiva and Lerman, 1985). In the past decades, Multinomial Logit (MNL) model is the simplest and most practical one in route choice problems within the RUM framework (Prato, 2009). The classical MNL model assumes that error terms are identically and independently distributed (I.I.D.). However, this assumption may result in biased estimation due to their partial
overlap of alternative routes (Bovy et al., 2008). To deal with the overlapping problem, two main approaches were proposed, i.e., incorporating a correction term to approximate the correlation among alternative within the MNL structure, and constructing a new structure such as Generalized Extreme Value (GEV) model. Cascetta et al. (1996) proposed the C-logit model, in which the degree of similarity of each alternative with other alternatives was represented by a commonality factor. Ben-Akiva and Bierlaire (1999) introduced the Path-size Logit (PSL) model, in which the path size indicates the fraction of the path that constitutes a full alternative. A unique path has a size equal to one and N duplicate paths share the size of 1/N. Besides the modifications of MNL, GEV models are also popular for overcoming the overlapping problem. GEV models account for similarities within the stochastic part of the utility function. Some widely used models of GEV in route choice problems are the link-based crossed nested Logit (CNL) (Vovsha and Bekhor, 1998; Lai and Bierlaire, 2015), paired combinatorial Logit (PCL) (Prashker and Bekhor, 1998), and Generalized Nested Logit model (Bekhor and Prashker, 2001). These models all have a nest structure to present the link-path relation, in which alternatives with shared attributes are grouped into the same nest so the correlation can be explicitly captured. However, GEV models are usually not easy to get feasible results in a large-scale network due to their structural complexity and computation burden. From a practical viewpoint, Bovy et al. (2008) pointed out that using the simple PSL model was a suitable alternative to the complex models, even though GEV models might handle correlations more exactly.

Even though RUM based models such as C-logit and PSL are frequently used in practice due to their simple structure, several investigations have shown the violation of the utility-maximizing assumption (Allais and Hagen, 1979; Avineri and Prashker, 2004; Bogers et al., 2005). Alternatively, two possible approaches were proposed in the literatures: suggesting another structure to RUM based models and incorporating psychological and other factors to RUM based models. Recently, the RRM based approach to discrete choice modeling provides an alternative to the conventional linear-additive RUM based approach, which is built on the psychological notion that regret can be regarded as an important determinant of choice behavior (Chorus, 2010; Chorus, 2012; Chorus, 2014; van Cranenburgh et al., 2015). The regret can be conceptualized as the emotion experienced when one or more non-chosen alternatives perform better than a chosen one in terms of one or more attributes. The most concerned question is how it performs when compared with RUM based models. Several researches had estimated these two competing model paradigms on a number of datasets ranging from mode choices, car type choices, parking lot choices, and route choices (Chorus, 2010; Chorus and de Jong, 2011; Chorus et al., 2011; Hensher et al., 2013; Prato, 2014). A strong performance of RRM based models were shown in their empirical studies. For example, Prato (2014) found that RRM based models perform better than the RUM based models for most cases in route choice problem. In addition, some studies had developed hybrid models in which a subset of attributes is subject to RUM, while another subset is subject to RRM (Chorus et al., 2013; Hess and Stathopoulos, 2013), which could also improve the model performance.

It is widely accepted that individuals have varying tastes to specific attributes, and hence, choice results. The heterogeneity should be investigated underlying RUM or RRM based approach. Generally, it can be divided into observed and unobserved heterogeneity. In a reveal survey for route choice analysis, the observed heterogeneity (to an analyst) could include the OD pair specific heterogeneity such as OD distance, personal specific heterogeneity such as age and gender, and time specific heterogeneity such as peak/off-peak hours and day of week. The
unobserved heterogeneity (to an analyst) might include trip purpose and personal privacy such as income. To incorporate the observed heterogeneity to route choice modelling, Li et al. (2013) proposed two methods, i.e., structured scale parameter and structured parameters of explanatory variables. They found that the models incorporating OD familiarity fit the data better and their studies showed that trips between more familiar OD pairs had larger error variances and less sensitivity to route attributes. On the other hand, the mixed logit model (Hensher and Greene, 2003) is a popular mathematical structure for the analysis of unobserved heterogeneity. Bhat (2000) formulated a mixed logit model of travel mode choice that accommodated variations in mode preferences and responsiveness to level-of-service due to both observed and unobserved individual characteristics. The model parameters are estimated using a maximum simulated log-likelihood approach. Due to the difficulty of data collection for route choice analysis, we only found limited research considering observed individual heterogeneity with reveal preference data. Extensions to build in preference heterogeneity under RRM are also limited. As far as the authors aware, only one study has explicitly investigated the heterogeneity under the RRM framework (Hensher et al., 2016). They explored the difference between RUM and RRM when preference heterogeneity is account for through random parameters. They identified a statistically richer improvement in fit of mixed logit compared to multinomial logit under RRM and RUM but found small differences overall between the empirical outputs of RRM and RUM.

3. Methodology

In this section, we start with the introduction of the classical RRM model and Path-size RRM model. Next, we formulate the RRM based route choice model that allows for the representation of observed heterogeneity.

3.1 The classical RRM model

The concept of anticipated regret is an important determinant of choice behaviour. Regret is what one experiences when a non-chosen alternative performs better than the chosen one, and regret-based choice theories (Loomes and Sugden, 1982) and models are developed on the notion that individuals aim to minimize their perceived regret when making choices. Following the regret theory, random regret minimization (RRM) models (Chorus, 2010; Chorus, 2012; Chorus et al., 2013) postulate that decision makers try to minimize their regret when choosing alternatives. The level of perceived regret is associated with the considered alternative $i$. The regret of alternative $i$ is described by the sum of binary regrets where alternative $i$ is compared to other alternatives on each attribute in the personal choice set. This attribute-level regret can be formulated as follows.

$$ R_{i \to j}^m = ln \left( 1 + \exp \left( \beta_m (x_{j,m} - x_{i,m}) \right) \right) $$

(1)

where

- $R_{i \to j}^m$: the random regret associated with the considered alternative $i$ that is compared to alternative $j$ on attribute $m$.
- $\beta_m$: the estimated parameter associated with attribute $x_m$.
- $x_{i,m}, x_{j,m}$: the values associated with attribute $x_m$ for, respectively, the considered alternative $i$ and another alternative $j$.

This formulation implies that regret is close to zero when alternative $j$ performs worse than $i$ in terms of attribute $m$, and that it grows as an approximately linear function $\beta_m (x_{j,m} - x_{i,m})$ of
the difference in attribute values in the case when \(i\) performs worse than \(j\) in terms of attribute \(m\). Hence, the RRM based model postulates that when a decision maker considers alternative \(i\) as compared to alternative \(j\) he or she experiences almost no regret with regard to attribute \(m\) when the attribute \(m\) of alternative \(i\) performs considerably better (Chorus et al., 2014).

The overall regret is conceived to be the sum of all binary regrets associated with the binary comparisons between a considered alternative \(i\) and its competitor alternatives \(j\) for all attributes. The classical RRM model is proposed by Chorus (2010) as a smoothed approximation of

\[
\ln \sum \ln \sum = \ln \ln \sum \ln \sum = \sum \sum \sum \ldots
\]

and

\[
\sum \sum \sum \ln \sum \ln \sum = \sum \sum \sum \sum \ldots
\]

where

\[ R_i = \sum_{j \neq i} \sum_m \ln \{1 + \exp[\beta_m(x_{j,m} - x_{i,m})]\} \]  \( (2) \)

where

\( R_i \): the observed regret associated with alternative \(i\).

Similar to the RUM based framework, the functional form of the choice probabilities changes as different assumptions on the random error term \(\varepsilon_i\) are imposed. When the negative of the errors is assumed to be I.I.D. Type I Extreme Value, the choice probability can be derived using a classical MNL-formulated as follows.

\[
P_{n,i} = \frac{\exp(-R_i)}{\sum_{j \in C_n} \exp(-R_i)}
\]  \( (3) \)

Where

\( P_{n,i} \): the probability of selecting route \(i\) for individual \(n\);
\( C_n \): the choice set for individual \(n\).

### 3.2 The Path-size RRM

Because of the partial overlap of alternative routes in a route choice situation, the classical RRM model is not appropriate for route choice analysis. Prato (2014) expanded the paradigm of classical RRM in the route choice context considering similarities across alternatives. Three approaches are proposed: (1) adding utility-based corrections; (2) adding a regret-based term that compares the degree of similarity of a route with other alternatives; (3) adding a regret-based term that adjusts for the correlation of each route with any other alternatives. Empirical studies showed that the second and the third approaches performed slightly better than the first approach. Therefore, we follow the second approach to overcome the overlapping problem because of its simple formulation. In the second approach, a path size correction term (Ben-Akiva and Bierlaire, 1999) is added to the regret function, which expresses the pairwise comparison of the degree of independence of alternatives. The Path-size RRM (PS-RRM) can be formulated as follows.

\[
P_{n,i} = \frac{\exp(-\sum_{j \neq i} \sum_m \ln \{1 + \exp(\beta_m(x_{j,m} - x_{i,m}))\})}{\sum_{j \in C_n} \exp(-\sum_{k \neq j} \sum_m \ln \{1 + \exp(\beta_m(x_{k,m} - x_{i,m}))\})}
\]  \( (4) \)

\[
PS_i = \sum_{a \in I_i} \frac{\sum_{a \in I_i} \frac{L_a}{M_{a,C_n}}}{L_i}
\]  \( (5) \)

where

\( I_i \): the set of links in route \(i\);
\( \beta_{ps} \): the estimated parameter associated with path size;
\( L_a \): the length of link \(a\);
\( L_i \): the length of route \(i\).
$M_{a,c,n}$: the number of paths in $C_n$ using link $a$;

Intuitively, the expected sign of the parameter $\beta_{ps}$ is positive to indicate that the regret for the considered route decreases if this route is more independent than the alternative ones, and increases when this route is less distinct than the alternative ones.

### 3.3 The PS-RRM with heterogeneity

Individual’s responsiveness or taste to route attributes affects her or his route choice for a trip. This responsiveness will, in general, vary across individuals based on individual characteristics. The regret an individual associates with a chosen route can be viewed as comprising two components from the perspective of an analyst. The first component can be viewed as the intrinsic bias of the individual toward the route due to individual heterogeneity (e.g., age, gender, trip purpose, departure time). The second component can be viewed as the regret that the individual derived from perceived or observed attributes of route (e.g., travel time, distance, the number of signalized intersection, cost, and driving comfort). Therefore, we should obtain individual heterogeneity for the regret evaluation on route attributes.

In the PS-RRM model, we assumed that individuals are homogeneous and tastes $\beta_m$ are fixed coefficients. To incorporate individual heterogeneity, as applied by Bhat (2000) and Li et al. (2016) in RUM based models, $\beta_m$ are assumed to have a linear relationship between individual characteristics and route attribute coefficients. Therefore, $\beta_m$ in the PS-RRM model can be reformulated as follows.

$$
\beta_{n,m} = \sum_p \alpha_p y_{n,m,p} + \gamma_m
$$

where

- $y_{n,m,p}$: the observed variable $p$ that related to individual $n$’s tastes on route attribute $m$;
- $\alpha_p$: the estimated parameter associated with heterogeneity;
- $\gamma_m$: the estimated parameter associated with the constant term on route attribute $m$.

For route choice analysis, the individual heterogeneity can be divided to three categories: personal-specific (e.g. age and gender), OD pair specific (e.g. OD distance) and time specific (e.g. peak/off peak hours and day of week). We discuss the effect of heterogeneity on route attributes in section 5.

### 4 Data

The GPS and OBD data used in this study was collected from private vehicles. In recent years, benefiting from the popularity of vehicle navigation system, GPS data has become an important resource and has been used in route choice analysis and route-finding problems (Zeng et al., 2015; Zeng et al., 2016; Li et al., 2016). The data is collected from 150 private cars in Toyota city, Japan in 2011 as a part of the Green Mobility project supported by Toyota Motor Corporation. On-board equipment (e.g., OBD device) installed in their private cars recorded the GPS trajectory as well as the driving behaviour such as brake and acceleration operation second by second. The data are uploaded to the internet server by the participants every week. In this study, the GPS data can be used for travel time estimation and the driving behaviour data recorded by OBD can be used for driving comfort estimation.

#### 4.1 Road network

A road network with 4072 nodes and 12,877 links in Toyota city, Japan, is used to analyse the route choice behaviour. This is a dense network. It covers an area of about 320 km².
4.2 Observations

The 150 drivers who made trips every day in the period March to December of 2011 are selected as the subjects for this study. Because the data set is large-scale, we only select 4312 OD pairs in March for this empirical analysis. As shown in Fig. 1, the OD pairs with only one trip account for 69.29% of the total trips, while the OD pairs with 2-3 trips, 4-5 trips, 6-10 trips and 11 above trips only account for 14.49%, 5.59%, 2.94%, 7.70%, respectively. For the OD pairs with large number of trips (e.g., more than 10 trips), it is possible to extract the experienced routes to generate the choice set from the trip record if the driver used more than one routes. However, it is difficult to obtain the route choice set for most of the OD pairs with sparse trips. To fill this gap, it is necessary to generate the route choice set for each OD pair.

Because the travel time and driving comfort are considered in our route choice model, we need to estimate the route travel time and quantify the driving comfort for each candidate route. The average route travel time can be obtained by using GPS data after map-matching process (Miwa et al., 2012), and the driving comfort is quantified by heart rate increase which is estimated by easy-to-measured variables such as driving condition and driving operation collected by GPS and OBD.

![Fig. 1 The distribution of the number of trip for each OD pair](image)
4.3 Driving comfort estimation

Travel time, distance, cost, value of time, the number of signalized intersection were usually regarded as the important route attributes in route choice analysis. However, driving comfort has not been explicitly considered due to the difficulty in data collection, although it is an important influence factor when drivers choose the best routes. Some drivers may wish only to minimize travel time. Others may feel uncomfortable making difficult maneuvers, and therefore avoid lane changes, freeways, heavily-congested roads or left turns (right turn traffic) at intersections.
without protected signals (Ramming, 2001). Elder drivers might be more insensitive to travel time and distance, given the comfort and better driving conditions that they enjoy. Some studies have shown that driving comfort or stress can be indicated by physiological signals such as heart rate, respiration, muscle activity, and skin conductance (Healey and Picard, 2005; Yamakoshi et al., 2009; Zeng et al., 2017). As shown in Fig. 2, drivers adjust their operation to accommodate the change of driving environment. And the driving environment and operation will cause the change of driving comfort; meanwhile the driving comfort also reacts on the driving operation. Heterogeneous drivers are assumed to choose the best routes based on their feeling of driving comfort and travel time in the past driving experience. In this study, we use heart rate increase as an indicator for driver comfort because it is convenient to collect by portable device such as Polar monitor (Goodie et al., 2000; Giles et al., 2016). Driving comfort can be quantified by the heart rate increase compared with the heart rate in calm down situation. Large heart rate increase usually indicates nervous and stressful mental state when the driver needs to deal with complex situation. As shown in Fig. 3, the link-based driving comfort was detected in a ring route (Zeng et al., 2017). Totally 21 variables related to driving environment and driving operation, such as speed, acceleration, brake, vehicle confliction, pedestrian confliction, lane change, are collected for driving comfort analysis. Our previous study (Zeng et al., 2017) has applied machine learning approach such as Random Forest to estimate the driving comfort and found that the top four easy-to-measured variables, i.e., average link speed, standard deviation of link speed, average brake frequency and average acceleration frequency, make a contribution of 72% to the driving comfort. Therefore, it is possible to estimate the link-based driving comfort based on these four variables. One of the significant advantages is that these four variables can be easily collected by the in-vehicle devices such as OBD and GPS. Since these four variables can represent most of the traffic condition and driver operation, it is not necessary to install the camera and physiological devices to capture the traffic condition and driving comfort in a large-scale data collection procedure. Following our previous study, we apply the machine learning approach to estimate the link-based driving comfort in the whole network. And the route-based driving comfort can be represented by the sum of the link-based driving comfort after calculating the average link speed, standard deviation of link speed, average brake frequency and average acceleration frequency for each link.

### 4.4 Route choice set generation

Since the road network used in this study is a relatively large-scale network, the population of available routes for an OD pair (the universal set) is very large and mostly not known. The identification of distinct relevant route alternatives in such a network is not straightforward and requires model-based approaches such as repeated shortest path search (Bovy, 2009). Some studies used the Monte Carlo simulation for searching the shortest path with distance as criterion where link distances were repeatedly randomized using a normal distribution (Frejinger et al., 2009; Lai and Bierlaire, 2015; Li et al., 2016). However, a driver would not recognize the difference among the similar routes with highly partial overlap. Generally, the universal route set will not be known by the driver due to the large number of potential paths and his or her limited cognitive abilities. The driver cognition is usually associated with his or her travel experiences in this network and his or her manner of acquiring information. Therefore, we propose to generate the route choice set on the experienced road network but not on the whole network. As shown in Fig. 4, we extract the links that the driver experienced in the past one month to build the
individual network for generating the route choice set. This procedure can greatly reduce the number of possible routes and guarantee that all the routes in the experienced network can be cognized by the driver. Then, the route choice set is generated by using a link penalty method which is similar to Chen et al. (2007). As shown in Algorithm 1, we first generate the shortest route based on the average travel time, then the selected links are given a penalized weight and the next shortest route based on the penalty network will be generated. We iteratively penalty the selected links and generate k routes until the travel time of the last route exceed 1.5 times of the first shortest route. Finally, we add the observed route into the choice set if the observed route is not included. Fig. 5 gives an example of the individual route choice set.

Algorithm 1. Route choice set generation

1. **Step 1**: Initialization
   2. Set \( S_n = \) “empty” for all OD pairs \( n \);
   3. Set \( K_n = 1 \);
   4. Find the least travel time route \( P_{n, K_n} \) for all OD pairs with A-star algorithm;
   5. Compute the route travel time \( T_{n, K_n} \);
   6. Save \( P_{n, K_n} \) to \( S_n \);

2. **Step 2**: Iteration
   8. For each OD pair \( n \), set iteration number \( m = 0 \);
   
   9. **Step 2.1**: Link weight penalty
      10. For each link on a path in \( S_n \), their weights were penalized by \( w_i \), where
      11. \( w_i = \alpha^m W_i \) with \( 0 < \alpha < 1 \), and \( W_i \) is the original weight of link \( i \);
      12. \( K_n := K_n + 1 \);
      13. \( m := m + 1 \);

   14. **Step 2.2**: Find the candidate paths
       15. Calculate the shortest path \( P_{n, K_n} \) based on \( \alpha^m W_i + W_i \);

   16. **Step 2.3**: Path checking
      17. Compute the path travel time \( T_{n, K_n} \) based on the sum of \( W_i \) in the path \( P_{n, K_n} \)
      18. If \( T_{n, K_n} < 1.5 T_{n, 1} \) and \( P_{n, K_n} \notin S_n 
      19. \quad \text{Save } P_{n, K_n} \text{ to } S_n \);
      20. Else
         21. \( K_n := K_n - 1 \)

   22. **Step 2.4**: Termination
      23. If \( K_n < N \), where \( N \) is the upper number of candidate path
      24. \quad Go to **Step 2.1**;
      25. Else
         26. Add the observed path into \( S_n \)
         27. **End**;
Fig. 4 The whole network and personally experienced network

Fig. 5 The route choice set based on personally experienced network
5 Empirical results

5.1 Model specification

Two route attributes, average travel time and driving comfort are included into the regret function. Travel time and distance are two highly similar and correlated attributes, so only one of them would be sufficient to the regret function. The cost is not considered because almost all of the roads are free in the urban road network. Since driving comfort should be highly correlated with the number of intersections, signal control and road grade, these variables are not considered. To illustrate the performance of the PS-RRM model with heterogeneity (PS-RRMH), three other models, i.e., Path-size logit (PSL), Path-size logit with heterogeneity (PSLH), and PS-RRM, are also estimated in this analysis. A summary of the different structures is shown in Table 1. The systematic part of the regret with observed heterogeneity in Eq. (6) is given as:

$$\beta_{n,m} = \alpha_1 \text{DayOfWeek}_{n,m} + \alpha_2 \text{DepartureTime}_{n,m} + \alpha_3 \text{Young}_{n,m} + \alpha_4 \text{Old}_{n,m} + \alpha_5 \text{Gender}_{n,m} + \alpha_6 \ln(OD\_distance) + \gamma_m$$

The PSL model is given as follows.

$$P_{n,i} = \frac{\exp(\sum_m \beta_m x_{i,m} + \beta_{ps} \ln(PS_i))}{\sum_{j \in C_n} \exp(\sum_m \beta_m x_{j,m} + \beta_{ps} \ln(PS_j))}$$

(8)

Accordingly, the PSLH model is given as follows.

$$P_{n,i} = \frac{\exp(\sum_m \beta_{m,n} x_{i,m} + \beta_{ps} \ln(PS_i))}{\sum_{j \in C_n} \exp(\sum_m \beta_{m,n} x_{j,m} + \beta_{ps} \ln(PS_j))}$$

(9)

where $$\beta'_{n,m} = \beta_{n,m}$$.

5.2 Estimation result

The estimation results and goodness of fit of the four models are shown in Table 1. The Constant_T, Constant_C and Constant_P are fixed parts of the tastes on route attributes. The Constant_C and Constant_P have a negative sign as expected. However, Constant_T has a positive sign, which would be expected to have a negative sign because longer travel time should reduce the utility or increase the regret. To further explain the unusual estimation result, we estimate two PS-RRM models by considering travel time and driving comfort separately. As shown in Table 2, all of the parameters have expected signs. Interestingly, it is found that the likelihood of PS-RRM only considering driving comfort and path size is greatly better than the likelihood of PS-RRM only considering travel time and path size. That means the effect of driving comfort is much stronger than the effect of travel time on the model fitness. Therefore, we guess the unexpected sign of Constant_T is caused by the dominated effect of driving comfort.

Then, we look at the performance of the four models. As expected, the PSLH model and PS-RRMH model are better than their original form without considering the observed heterogeneity. The PS-RRMH model is better than the PSLH model, which confirms our assumption that drivers are more likely to choose routes by minimizing their anticipated regret instead of maximizing their utility.

The OD pair specific heterogeneity is reflected by the logarithm of OD distance. The negative sign and t-statistic suggest that OD distance will affect the taste for average travel time and driving comfort significantly. Drivers are more sensitive to average travel time and driving comfort when they need to drive longer distances. It is reasonable because drivers usually need
to make a travel time schedule for a long distance trip on one hand and they prefer to comfortable driving that prevents fatigue on the other hand.

The personal specific heterogeneity is represented by age and gender. All of these variables are significant. It is found that young drivers (age<=35) are less sensitive to travel time and driving comfort; while elder drivers (age>60) are less sensitive to travel time but they are more sensitive to driving comfort. It also indicates that the middle age drivers are more sensitive to travel time than the young and elder drivers, while they are less sensitive to driving comfort than elder drivers.

The time specific heterogeneity is represented by departure time (depart at peak/off-peak hours) and day of week (weekend/weekday). The negative sign of departure time indicates that drivers are more sensitive to travel time and driving comfort at peak hours, because the travel time may be unstable and the congestion and frequent traffic conflict may cause uncomfortable driving at peak hours. It is also found that people are more sensitive to travel time when they depart at weekend, while they are less sensitive to driving comfort.

The positive sign and t-statistic for the path size indicates that the regret will reduce and the utility will increase if the route has less overlap than the alternative ones.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PSL</th>
<th>PSLH</th>
<th>PS-RRM</th>
<th>PS-RRMH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>t-stat</td>
<td>Est</td>
<td>t-stat</td>
</tr>
<tr>
<td>Average travel time</td>
<td>0.0081</td>
<td>25.55</td>
<td>0.0067</td>
<td>3.44</td>
</tr>
<tr>
<td>Constant_T</td>
<td>-0.000027</td>
<td>-2.04</td>
<td>-0.00052</td>
<td>-5.60</td>
</tr>
<tr>
<td>Day of week (1, weekend; 0, weekday)</td>
<td>-0.0024</td>
<td>-3.71</td>
<td>-0.00040</td>
<td>-4.59</td>
</tr>
<tr>
<td>Departure time (1, peak hour; 0, off-peak hour)</td>
<td>0.00020</td>
<td>2.02</td>
<td>0.00042</td>
<td>2.99</td>
</tr>
<tr>
<td>Young (1, age&lt;=35; 0, age&gt;35)</td>
<td>0.0027</td>
<td>2.70</td>
<td>0.00033</td>
<td>2.00</td>
</tr>
<tr>
<td>Old (1, age&lt;=60; 0, age&gt;60)</td>
<td>-0.00167</td>
<td>-2.07</td>
<td>-0.0011</td>
<td>-3.7</td>
</tr>
<tr>
<td>Gender (1, male; 0, female)</td>
<td>-0.00127</td>
<td>-2.13</td>
<td>-0.0018</td>
<td>-2.05</td>
</tr>
<tr>
<td>Logarithm of OD distance</td>
<td>-0.26</td>
<td>-39.51</td>
<td>-0.17</td>
<td>-3.27</td>
</tr>
<tr>
<td>Driving comfort</td>
<td>0.056</td>
<td>4.08</td>
<td>0.051</td>
<td>15.92</td>
</tr>
<tr>
<td>Constant_C</td>
<td>0.040</td>
<td>3.03</td>
<td>0.011</td>
<td>3.80</td>
</tr>
<tr>
<td>Day of week (1, weekend; 0, weekday)</td>
<td>0.0025</td>
<td>2.14</td>
<td>0.018</td>
<td>3.37</td>
</tr>
<tr>
<td>Departure time (1, peak hour; 0, off-peak hour)</td>
<td>-0.0027</td>
<td>-2.12</td>
<td>-0.037</td>
<td>-4.32</td>
</tr>
<tr>
<td>Young (1, age&lt;=35; 0, age&gt;35)</td>
<td>0.037</td>
<td>2.26</td>
<td>0.060</td>
<td>4.69</td>
</tr>
<tr>
<td>Old (1, age&lt;=60; 0, age&gt;60)</td>
<td>-0.046</td>
<td>-7.42</td>
<td>-0.046</td>
<td>-10.65</td>
</tr>
<tr>
<td>Gender (1, male; 0, female)</td>
<td>14.74</td>
<td>20.50</td>
<td>14.85</td>
<td>15.24</td>
</tr>
</tbody>
</table>
Sample size 4312
Null LL -8482.39
Final LL -2956.31 -2908.55 -2868.46 -2759.13
Rho_sq 0.651 0.657 0.662 0.675

<table>
<thead>
<tr>
<th>Path Attributes</th>
<th>PS-RRM (with driving comfort and path size)</th>
<th>PS-RRM (with travel time and path size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel time</td>
<td>--</td>
<td>-0.00047 -21.33</td>
</tr>
<tr>
<td>Driving comfort</td>
<td>-0.057</td>
<td>-41.11</td>
</tr>
<tr>
<td>Path size</td>
<td>2.10</td>
<td>28.98 1.75 31.19</td>
</tr>
<tr>
<td>Sample size</td>
<td>4312</td>
<td></td>
</tr>
<tr>
<td>Null LL</td>
<td>-8482.39</td>
<td></td>
</tr>
<tr>
<td>Final LL</td>
<td>-3461.08</td>
<td>-7578.29</td>
</tr>
<tr>
<td>Rho_sq</td>
<td>0.592</td>
<td>0.107</td>
</tr>
</tbody>
</table>

6 Conclusions

This study developed a route choice model based on RRM approach. In contrast to RUM approach, RRM enables to minimize the anticipated regret of not having chosen another route when making choices. Comparing with RUM based models, RRM based models show better fitness, which indicates the assumption that drivers choosing the best route by minimizing their anticipated regret are more appropriate, even though the assumption of maximizing the user utility is still reasonable.

The first contribution of this study is the application of PS-RRM model that explicitly captures the heterogeneity which includes the OD pair specific heterogeneity, personal specific heterogeneity, and time specific heterogeneity. Second, the driving comfort represented by heart rate increase is explicitly incorporated into the route attributes. Previous studies usually regarded driving comfort as an unobserved attribute, even though it should be regarded as an important factor to route choice. We estimate the driving comfort by using four easy-to-measured variables which can be captured by OBD and GPS. The estimation results confirm that the incorporation of observed heterogeneities such as OD distance, age, gender, departure time and day of week enables to improve the performance of the RRM and RUM based models. The t-statistic indicates that the observed heterogeneities have significant effects on taste of travel time and driving comfort. For example, drivers are more sensitive to average travel time and driving comfort when they need to drive longer distances. The middle age drivers are more sensitive to travel time than the young and elder drivers, while elder drivers show more sensitive to driving comfort than young and middle age drivers. The time specific heterogeneity shows that drivers are more sensitive to both travel time and driving comfort at peak hours and weekend, while they are more sensitive to travel time but less sensitive to driving comfort.

In further research, the GPS and OBD data might be combined with questionnaire data so as to take into account for a greater number of behaviour terms such as activity schedule and travel time budget. Further, in this study, the unobserved heterogeneity is not incorporated so as not to make the model too complicated. However, a mixed RRM based model should be formulated to capture the random effects of the unobserved heterogeneity.
References
Chen, Y., Bell, M. G., & Bogenberger, K. (2007). Reliable pretrip multipath planning and
dynamic adaptation for a centralized road navigation system. IEEE Transactions on Intelligent Transportation Systems, 8(1), 14-20.


process: A comparison of planned and observed routes obtained using person-based GPS. Transportation research part F: traffic psychology and behavior, 12(4), 347-358.