EJTIR

Exploring travellers' risk preferences with regard to travel time reliability on the basis of GPS trip records

Weiliang Zeng[1]
Institution of Materials and Systems for Sustainability, Nagoya University.

Tomio Miwa[2]
Institution of Materials and Systems for Sustainability, Nagoya University.

Takayuki Morikawa[3]
Institute of Innovation for Future Society, Nagoya University.

Travel time reliability has attracted considerable interest in the field of route choice modelling. Knowing how individuals choose paths with uncertain travel times is fundamental to advancing our understanding of route choice behaviour and thus driving the development of route guidance systems. In general, existing navigation systems provide the shortest path on the basis of distance or travel time, even though many travellers do not intend to choose the shortest path. Several studies have shown that the probability of delay or travel time reliability is an important factor in a traveller's route choice decision. Learning a traveller's risk preference with regard to travel time reliability is important for designing a preferable route. Traditionally, route choice data for individual preference analysis are collected by conducting stated preference surveys. However, this approach is difficult to avoid its inherent limitation, namely a lack of honest, accurate, and bias-free reporting. To overcome these problems, the present study proposes a new data collection methodology that facilitates estimation of a traveller's risk preference on the basis of large-scale GPS trip records. The lower and upper bounds of individual risk preference can be estimated by exhausting a series of reliable paths with different on-time arrival probabilities and using the theory of stochastic dominance. Then, a regression model based on a logistic function is established to explore how socio-demographic and trip characteristics influence the lower and upper bounds. Thus, individual properties, such as age, and pre-trip information, such origin-destination (OD) distance, departure time, and day of week, are found to have a significant influence on the degree of risk preference.

Keywords: Risk preference, travel time reliability, stochastic dominance, reliable path, GPS probe vehicle.

1. Introduction

Car navigation systems have witnessed rapid development and widespread use in recent years. As a result, in addition to the travel time, distance, and emission, the on-time arrival probability or reliability is also regarded as an important route search criterion (Zeng et al., 2015; Zeng et al., 2016a; Zeng et al., 2016b). To develop an effective trip plan, a reliable path with a given on-time arrival probability is found to be more attractive than the shortest path in a stochastic network

[1] A: Furo-cho, Chikusa, Nagoya, Japan 464-8603 T: +81-52-789-3730 F: +81-52-789-5728 E: zengweiliang@imass.nagoya-u.ac.jp
[2] A: Furo-cho, Chikusa, Nagoya, Japan 464-8603 T: +81-52-789-5818 F: +81-52-789-5728 E: miwa@nagoya-u.jp
[3] A: Furo-cho, Chikusa, Nagoya, Japan 464-8603 T: +81-52-789-3564 F: +81-52-789-5728 E: morikawa@nagoya-u.jp

(Xing and Zhou, 2011). Travellers may exhibit different route choice behaviours depending on their risk preferences with regard to the stochastic travel time. Risk-averse travellers are likely to choose a more reliable path with a higher on-time arrival probability within their travel time budget, while risk-taking travellers are likely to prefer the shortest path that potentially involves the shortest distance or minimum travel time (Ben-Elia et al., 2013).

Socio-demographic characteristics and trip characteristics have been found to influence the risk preference. For example, de Palma and Picard (2005), who studied route choice behaviour under uncertain travel time, found that risk aversion was greater for transit users, blue-collar workers, and those scheduled for business appointments. A value pricing experiment for a tolled route in California, USA, showed that the travel time reliability is influenced by gender, wage, time of day, and car occupancy (Lam and Small, 2001). Carrion and Levinson (2013) estimated the value of the improvement in travel time reliability in high-occupancy toll lanes. Heterogeneous travel factors, such as gender, habit, time of day, and travel time variability, were found to be statistically significant. Prato et al. (2014) calculated the value of congestion and the value of reliability using a large-scale GPS dataset. Their empirical results indicated that these values were significantly higher during peak hours because of the possibly higher penalties for being late. Small et al. (2005) employed a mixed logit model to study the distribution of commuters' preferences for fast and reliable highway travel. Their study indicated that travellers exhibited substantial heterogeneity in terms of their preferences of travel time reliability.

Route choice data for traveller preference analysis can be collected by conducting revealed preference (RP) or stated preference (SP) surveys. Abdel-Aty et al. (1995) conducted an SP survey to investigate the effect of travel time variability on route choice. Their results showed the significance of the degree of travel time reliability on the decision-making process for route selection. Small et al. (2005) identified the variable nature of travel preferences in terms of both travel time and reliability by conducting both SP and RP surveys on observed commuter behaviour. However, it is difficult to avoid the inherent limitation of such approaches, namely the lack of honest, accurate, and bias-free reporting. For example, respondents have to assume a route choice instead of experiencing it practically in the case of an SP survey. Thus, they may simply answer questions about a choice that that they would not make realistically. Although an RP survey can reflect decision-making preferences more realistically, few studies have used RP data to investigate travel time reliability because it is difficult to collect a sufficient number of real examples at the level of detail required for ascertaining the reliability estimates (Carrion et al., 2012). On the other hand, data collection and analysis for an SP survey is time-consuming and expensive. To overcome these drawbacks, the present study proposes a new data collection methodology that facilitates estimation of a traveller's risk preference on the basis of large-scale GPS-based trip records. Probe vehicles with GPS devices can effectively provide detailed travel information on the start and end time, precise route observation, and OD information. In recent years, many studies have used probe data to analyse route choice behaviour. Li et al. (2005) showed how GPS facilitates effective recording of observed route choice information by monitoring 182 travellers over a 10-day period. Papinski et al. (2009) explored the decision-making process of route choice by comparing the observed and planned routes obtained from personal GPS data. Papinski and Scott (2011) developed a GIS-based toolkit for route choice analysis. Lima et al. (2016) analysed frequently chosen routes and presented a spatial probability distribution that bounded the route selection space with an ellipse. However, the above-mentioned researchers did not analyse the risk preference with regard to travel time reliability.

In general, existing studies related to the reliable path problem (Sivakumar et al., 1994; Sen et al., 2001; Xing and Zhou, 2011; Chen et al., 2016) assume traveller homogeneity. The variable nature of travellers' risk preference is seldom taken into account. To alleviate this problem, Chen and Ji (2005) proposed the concept of α-reliable path, which allows travellers to specify a confidence level α for finding a reliable path with the minimum travel time budget. The confidence level α can be regarded as a surrogate index of the traveller's risk preference with regard to travel time

reliability. However, the α-reliable path definition requires travellers to express their expected risk preference in terms of travel time uncertainty (Chen et al, 2013; Zeng et al., 2015). Travellers may be confused when defining a suitable confidence level α prior to their trips without any reference or default values provided by the navigation system. A reasonable approach for improving automatic routing guidance is to incorporate travellers' risk preferences into the routing process. Therefore, it is necessary to investigate travellers' risk preferences with regard to travel time reliability on the basis of long-term trip records.

In summary, route choice behaviour analysis using emerging tools, such as GPS probe vehicles, is attracting increasing interest. However, studies that discuss how to quantify travellers' risk preferences are limited. The measure of risk preference with regard to travel time reliability and its application to the reliable path finding problem are relatively new concepts in the field of route choice behaviour analysis. The contributions of this study are twofold. First, we propose a new method for measuring the degree of risk preference with regard to travel time reliability on the basis of large-scale GPS trip records. Second, we develop a regression model to interpret how the socio-demographic characteristics and trip characteristics influence the risk preference.

2. Methodology

2.1 Assumption and problem statement

This study assumes that a rational traveller chooses a reliable path on the basis of his or her risk attitude. A risk-taking traveller may choose a path that potentially involves the minimum travel time, whereas a risk-averse traveller prefers a path with higher on-time arrival probability within the travel time budget.

Definition 1: The degree of traveller's risk preference (DTRP) with regard to travel time is defined as the value of the on-time arrival probability of the selected path within a specified travel time budget.

As shown in Figure 1, assume that the travel time budget is 50 min and the DTRP is 0.8 or 0.7 depending on whether Path 1 or Path 2 is chosen, respectively. Note that the DTRP is related to the travel time distribution and the specified travel time budget; it is necessary to understand the path travel time distribution and the individual travel time budget. The distribution of the path travel time can be estimated from the observed dataset. However, it is difficult to determine the travel time budget precisely from the observed trip travel time, because the travel time budget is usually larger than the actual travel time but it cannot be derived without the traveller's input. Thus, the estimation of DTRP is an intractable problem. To address this issue, we propose a new method for estimating the DTRP by exhausting a series of reliable paths with different risk preferences (on-time arrival probabilities) and using the stochastic dominance theory to bind the range of risk preference level.
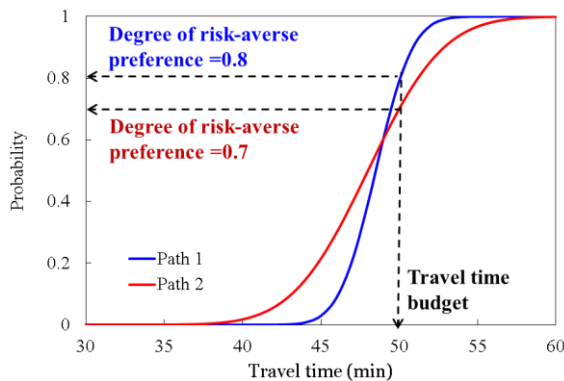

Figure 1. Definition of degree of traveller's risk preference

2.2 Path stochastic dominance

Given an on-time arrival probability α, a path $p_{rs}^* \in P_{rs}$ is defined as the α-reliable path if $\Phi_{T_{rs}^*}^{-1}(\alpha) < \Phi_{T_{rs}}^{-1}(\alpha)$ for any other path $p_{rs} \in P_{rs}$, where P_{rs} is the path set from the origin r to the destination s (Chen et al., 2013). Further, $\Phi_{T_{rs}^*}^{-1}(\alpha)$ and $\Phi_{T_{rs}}^{-1}(\alpha)$ are the inverse cumulative distribution functions (CDFs) of travel time for paths p_{rs}^* and p_{rs}, respectively. The α-reliable path problem has been solved by Lagrangian relaxation with Cholesky decomposition in our previous study (Zeng et al., 2015).

Definition 2 (FSD, first-order stochastic dominance): Given two paths $p_{rs}^a \neq p_{rs}^b \in P_{rs}$ where r is the origin and s is the destination, p_{rs}^a dominates p_{rs}^b (denoted by $p_{rs}^a \succ p_{rs}^b$) if $\Phi_{T_{rs}^a}^{-1}(\alpha) < \Phi_{T_{rs}^b}^{-1}(\alpha)$ for any on-time arrival probability $0 < \alpha < 1$.

Assume that the path travel time follows a normal distribution; then, the path stochastic dominance can be expressed by the mean and variance of the path travel time as follows.

Definition 3 (MVD, mean-variance dominance): Given an on-time arrival probability α and two paths ($p_{rs}^a \neq p_{rs}^b \in P_{rs}$), $p_{rs}^a \succ p_{rs}^b$ if p_{rs}^a and p_{rs}^b satisfy one of the following conditions:

(1) $u_a \leq u_b$ and $Z_\alpha \sigma_a < Z_\alpha \sigma_b$ or

(2) $u_a < u_b$ and $Z_\alpha \sigma_a \leq Z_\alpha \sigma_b$,

where

u_a, u_b: path travel time mean for p_{rs}^a and p_{rs}^b, respectively;

σ_a, σ_b: standard deviation of path travel time for p_{rs}^a and p_{rs}^b, respectively;

Z_α: inverse CDF of the standard normal distribution at α confidence level.

Proof: See Proposition 2 in Chen et al. (2013).

Figure 2 shows an example of path stochastic dominance based on travel time. Path 1, Path 2, and Path 3 dominate Path 4 for any on-time arrival probability $0 < \alpha < 1$. Therefore, Path 4 is the dominated path (unattractive path). Path 1 is not dominated by the other paths if the on-time arrival probability is greater than 0.12. Similarly, Path 3 is not dominated by the other paths if the on-time arrival probability is less than 0.12. Therefore, Path 1 and Path 3 are non-dominated paths. Now, let us consider the case of Path 2. According to FSD, Path 2 dominates Path 1 if $\alpha < 0.09$, and Path 2 dominates Path 3 if $\alpha > 0.16$. Thus, Path 2 is a non-dominated path with respect to Path 1 or Path 3. However, it is found that Path 2 is dominated by Path 1 if $\alpha > 0.09$, and it is dominated by Path 3 if $\alpha < 0.16$. Thus, Path 2 is dominated by Path 1 or Path 3 in the entire range of the on-time arrival probability. Therefore, Path 2 is a dominated path (unattractive path). A rational traveller will avoid Path 2 regardless of the degree of his or her risk preference. Both FSD and MVD can be used to identify the non-dominated path (attractive path) between two paths, but it is easier to use MVD to identify the non-dominated path by comparing the mean and standard deviation values of the travel times for the path set.

Definition 4 (Non-dominated path): A path $p_{rs}^a \in \mathbf{P_{rs}}$ is a non-dominated path if and only if p_{rs}^a dominates all paths $\forall p_{rs}^i \in \mathbf{P_{rs}}$ in a certain range of on-time arrival probability $0 < \lambda_1 < \alpha < \lambda_2 < 1$.

Here, we use **Definition 4** to judge whether the observed path is a non-dominated path in a path set. Since the α-reliable paths are the non-dominated paths between the OD nodes according to **Definition 4**, we do not need to generate all the paths between the OD nodes. Instead, the α-reliable paths can be used as the path set (p_{rs}^α). Then, MVD is used to judge whether the observed path is dominated by the α-reliable paths.
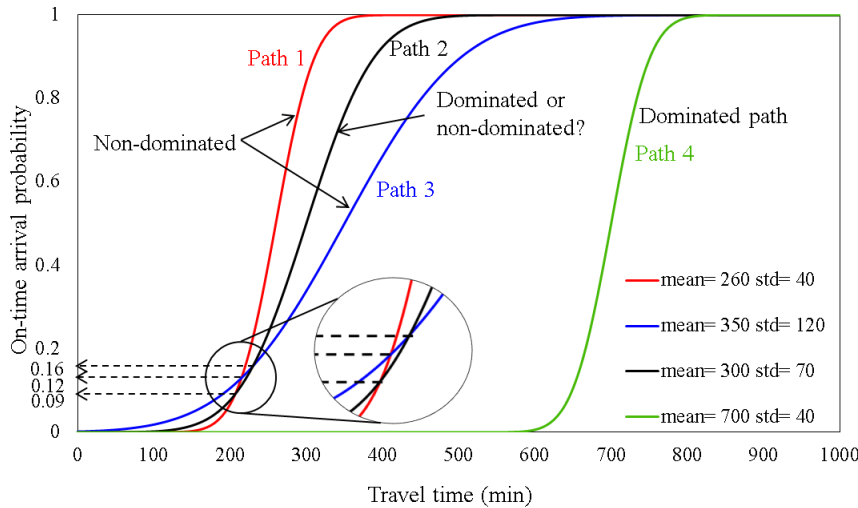
Figure 2. Path stochastic dominance based on travel time

First, we check whether the observed path p_{rs}^{obs} (with mean travel time u_{obs} and standard deviation of travel time σ_{obs}) is dominated by any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in p_{rs}^{\alpha}$) in the risk condition (α > 0.5). Since α > 0.5, $Z_{\alpha} > 0$, we need to check whether any existing α-reliable paths (with mean travel time u_{α} and standard deviation of travel time σ_{α}) satisfy (1) $u_{\alpha} \leq u_{obs}$ and $\sigma_{\alpha} < \sigma_{obs}$ or (2) $u_{\alpha} < u_{obs}$ and $\sigma_{\alpha} \leq \sigma_{obs}$. If any of the two conditions is satisfied, the observed path is regarded as the dominated path in the risk-averse condition.

Second, we check whether the observed path p_{rs}^{obs} is dominated by any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in p_{rs}^{\alpha}$) in the risk-taking condition (α < 0.5). Since α < 0.5, $Z_{\alpha} < 0$, we need to check whether any existing α-reliable paths satisfy (1) $u_{\alpha} \leq u_{obs}$ and $\sigma_{\alpha} > \sigma_{obs}$ or (2) $u_{\alpha} < u_{obs}$ and $\sigma_{\alpha} \geq \sigma_{obs}$. If any of the two conditions is satisfied, the observed path is regarded as the dominated path in the risk-taking condition.

According to MVD, the observed path is a dominated path if it is dominated by any of the α-reliable paths in the risk-averse or risk-taking condition. Thus, the observed path is the non-dominated path if it is dominated by no α-reliable path in the risk-averse or risk-taking condition. Based on the above discussion, the pseudocode for determining the non-dominated path is presented in Algorithm 1 in Table 1.

Table 1. Algorithm 1 for identification of non-dominated path

1	Step 1: Generate the α-reliable path set (Zeng et al., 2015): $\mathbf{p}_{rs}^{\alpha}$
2	Step 2: Check whether the observed path (p_{rs}^{obs}) is dominated in the risk-averse condition
3	Check MVD for any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in \mathbf{p}_{rs}^{\alpha}$):
4	(1) $u_{\alpha} \leq u_{obs}$ and $\sigma_{\alpha} < \sigma_{obs}$ or (2) $u_{\alpha} < u_{obs}$ and $\sigma_{\alpha} \leq \sigma_{obs}$
5	Step 3: Check whether the observed path (p_{rs}^{obs}) is dominated in the risk-taking condition
6	Check MVD for any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in \mathbf{p}_{rs}^{\alpha}$):
7	(1) $u_{\alpha} \leq u_{obs}$ and $\sigma_{\alpha} > \sigma_{obs}$ or (2) $u_{\alpha} < u_{obs}$ and $\sigma_{\alpha} \geq \sigma_{obs}$
8	Step 4: Determine the non-dominated path
9	Dominated path: Step 2 or Step 3 is satisfied
10	Non-dominated path: Step 2 and Step 3 are not satisfied

Assuming that travellers are rational (i.e., they never choose the dominated paths), we extract the non-dominated observed paths by using Algorithm 1. The dominated observed paths will be excluded for the estimation of the degree of risk preference. Because the α-reliable paths and the observed path are non-dominated paths, the cross between the CDF curves of the non-dominated

observed path and each α-reliable path must exist. Further, the value of the cross point determines the bound of the dominance condition between the two tested paths. The bound value of the degree of risk preference can be estimated by checking the cross points. The degree of risk preference (λ) for each cross point can be formulated as

$$\lambda = \Phi\left(\frac{u_{obs}\sigma_\alpha - u_\alpha\sigma_{obs}}{u_{obs}(\sigma_\alpha - \sigma_{obs})} - 1\right) \tag{1}$$

where

u_α, u_{obs}: path travel time mean for the α-reliable path and the observed path, respectively;

σ_α, σ_{obs}: standard deviation of path travel time for the α-reliable path and observed path, respectively;

Φ: CDF of the standard normal distribution.

The lower bound (λ_{LB}) and upper bound (λ_{UB}) of the degree of risk preference can be determined by comparing the observed path with the α-reliable paths. Obviously, the observed path dominates the α-reliable path with larger travel time variance (i.e., $\sigma_\alpha > \sigma_{obs}$) if $\alpha' > \lambda$, where α' is the desired risk preference. Therefore, the maximum value of λ can be regarded as the lower bound of the observed degree of risk preference.

$$\lambda_{LB} = \text{Max}\left\{\Phi\left(\frac{u_{obs}\sigma_\alpha - u_\alpha\sigma_{obs}}{u_{obs}(\sigma_\alpha - \sigma_{obs})} - 1\right)\right\}, \text{if } \sigma_\alpha > \sigma_{obs}. \tag{2}$$

Similarly, the observed path dominates the α-reliable path with smaller travel time variance (i.e., $\sigma_\alpha < \sigma_{obs}$) if $\alpha' < \lambda$. Therefore, the minimum value of λ can be regarded as the upper bound of the observed degree of risk preference.

$$\lambda_{UB} = \text{Min}\left\{\Phi\left(\frac{u_{obs}\sigma_\alpha - u_\alpha\sigma_{obs}}{u_{obs}(\sigma_\alpha - \sigma_{obs})} - 1\right)\right\}, \text{if } \sigma_\alpha < \sigma_{obs}. \tag{3}$$

The observed path dominates all of the α-reliable paths if the desired risk preference is set as $\lambda_{LB} < \alpha' < \lambda_{UB}$. An example of approximating the degree of traveller's risk preference is shown in Figure 3. To estimate the traveller's risk preference from the observed path, we first calculate cross points with the four α-reliable paths. Because the travel time variances of Path 1, Path 2, and Path 3 are larger than that of the observed path, the lower bound of the traveller's risk preference is selected from among the cross points between the CDF curves of the observed path and these three α-reliable paths. Because the cross point between the CDF curves of the observed path and Path 1 has a maximum value of λ, i.e., $\lambda = 0.63$, the lower bound of the traveller's risk preference is 0.63. Because only Path 4 has a smaller travel time variance than the observed one, the cross point between the CDF curves of the observed path and Path 4 is used to estimate the upper bound of the traveller's risk preference, i.e., $\lambda_{UB} = 0.91$.
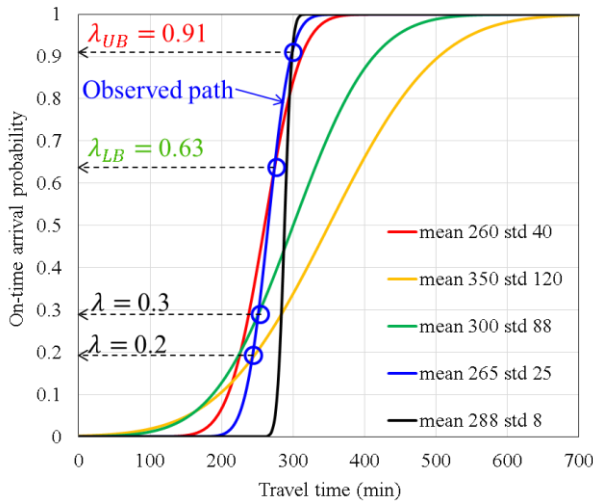
Figure 3. Estimation of traveller's risk preference

2.3 Modelling traveller's risk preference

Travellers' risk attitudes towards the stochastic travel time are important from the viewpoint of the reliable routing problem. To develop an intelligent navigation system, it is necessary to learn the travellers' risk preferences from their trip records. Here, a regression model is developed to estimate the effects of the explanatory variables on the travellers' risk preference. Because the degrees of risk preference are restricted to (0, 1), a logistic function is used to develop the regression model.

$$\lambda_{LB} = \frac{exp(\boldsymbol{\alpha}x+\varepsilon)}{1+exp(\boldsymbol{\alpha}x+\varepsilon)}, \tag{4}$$

$$\lambda_{UB} = \frac{exp(\boldsymbol{\beta}x+\varepsilon)}{1+exp(\boldsymbol{\beta}x+\varepsilon)}, \tag{5}$$

where $\boldsymbol{x}$ denotes the vector of observed explanatory variables, which include OD distance, age, departure time, day of week and gender; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the parameter vectors; and ε is the random error term, which is assumed to follow a normal distribution with zero mean and unit variance.

The logistic function can be reformulated in linear form as follows:

$$ln\left(\frac{\lambda_{LB}}{1-\lambda_{LB}}\right) = \boldsymbol{\alpha}x + \varepsilon, \tag{6}$$

$$ln\left(\frac{\lambda_{UB}}{1-\lambda_{UB}}\right) = \boldsymbol{\beta}x + \varepsilon. \tag{7}$$

3. Data and results

3.1 Road network

As shown in Figure 4, a road network with 4072 nodes and 12,877 links in Toyota City, Japan, was used to analyse the risk preference. This network covers an area of around 320 km2. The GPS data used in this study were collected from 153 probe vehicles in Toyota City, Japan. After map-matching (Miwa et al., 2012) and basic data cleaning, 3777 trip records were obtained for one month (March 2011).

3.2 Route set generation

The observed lower and upper bounds of risk preference were collected according to the proposed methodology. To generate sufficient reliable paths with various reliability levels, the range of α was set to 0.01–0.99 with intervals of 0.05. The route set was generated on the basis of the α-reliable path algorithm (Zeng et al., 2015). The α-reliable path problem can be formulated as the mean-variance problem. Cholesky decomposition and Lagrangian relaxation can be applied

to handle this problem, by which an intractable problem, with a non-linear and non-additive structure, can be decomposed into several tractable problems.

Figure 4. Road network

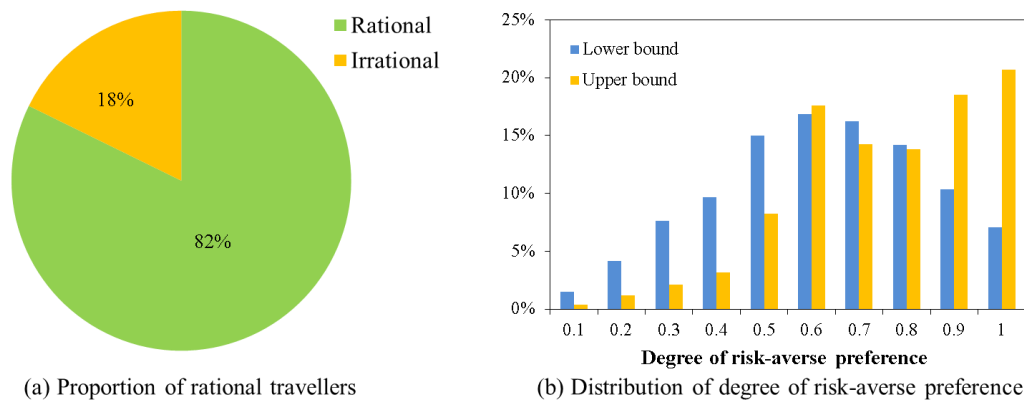(a) Proportion of rational travellers   (b) Distribution of degree of risk-averse preference

Figure 5. Estimation result of overall risk preference

3.3 Empirical analysis

As shown in Figure 5(a), 82% of the travellers were rational, and they selected the non-dominated paths. This implies that most of the local travellers could estimate the mean and variance of the path travel time because they were familiar with the traffic conditions and the road network. Figure 5 (b) shows the overall distribution of the lower and upper bounds of risk preference for all the trips. The average lower bound was 0.53 and the average upper bound was 0.71, which indicates that most of the travellers preferred reliable routes.

Figure 6 shows the average value of the lower and upper bounds of the degree of risk preference for different factors. It can be seen in Figure 6 (a) that as the OD distance increases, the proportion of high risk-averse preference increases. This implies that travellers have a higher risk-averse preference when they plan a longer trip. Figure 6 (b) shows that 54% of the elderly (age > 60 years) prefer highly risk-averse routes, while only 40% of the young (30 years < age < 40 years) prefer such routes. Around 6% of the highly risk-taking routes are chosen by people aged between 30 and 40 years, while only 3% of the highly risk-taking routes are chosen by the elderly. However, minimal risk aversion does not occur in the youngest age group (age < 30 years). One possible reason is that the steady travel time of the reliable path could result in less driving stress for elderly travellers. By contrast, young travellers aged between 30 and 40 years are more concerned with the minimum travel time because of commute or business concerns. Figure 6 (c) shows the risk attitudes for different departure times. It is found that people are more risk-averse during peak hours (8:00–10:00 and 18:00–20:00). The risk-averse attitudes decrease and the risk-

taking attitudes increase outside peak hours. Figure 6 (d) shows that more people prefer risk-averse routes on weekdays. Figure 6 (e) shows that female travellers are more risk-averse than male travellers.
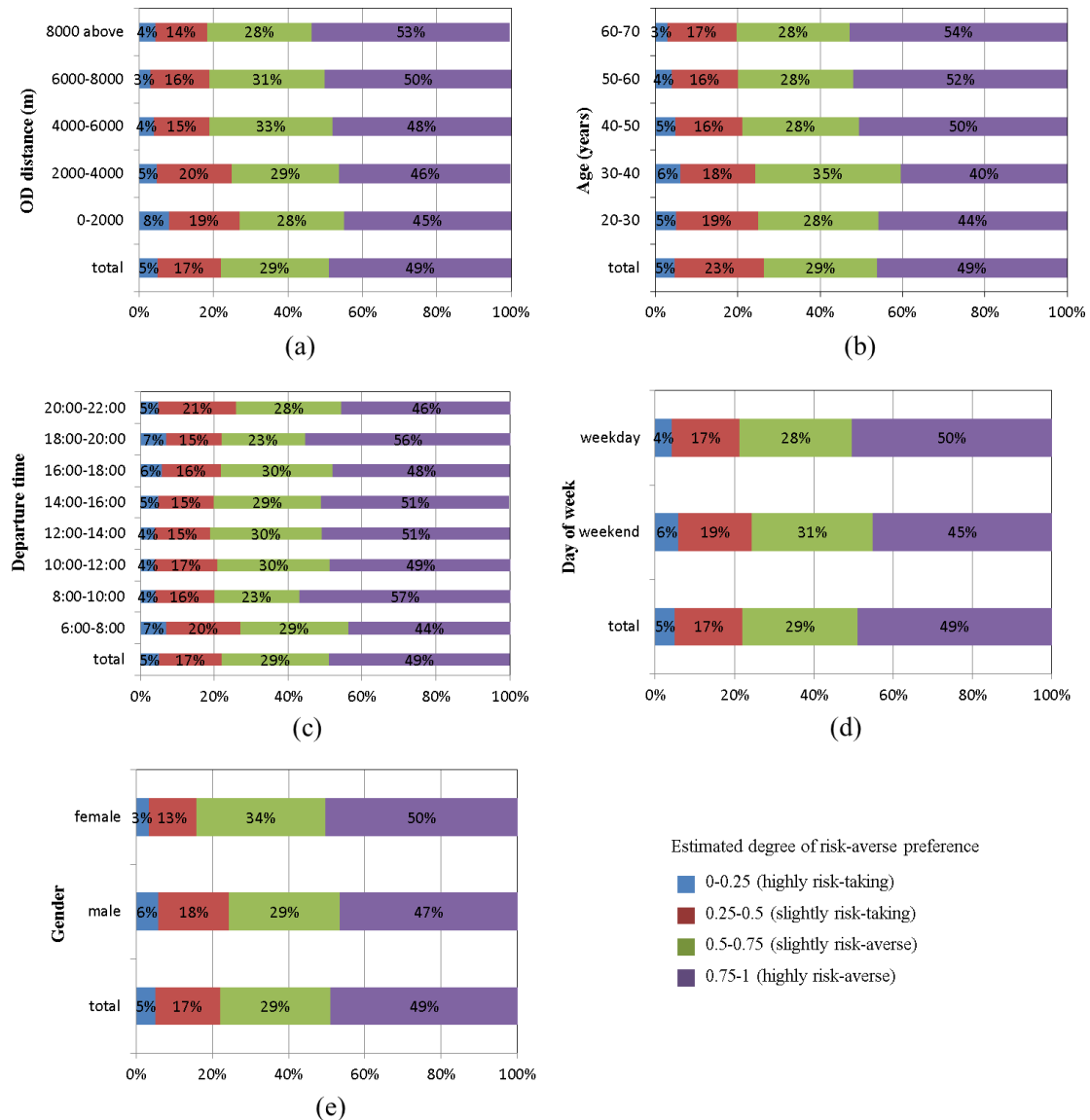

*Figure 6. Statistic of degree of risk preference for different factors*

3.4 Model estimation based on socio-demographic and trip characteristics

We estimated the parameters in the regression models by using the least-squares method. The explanatory variables include OD distance, age, departure time, day of week, and gender. It should be noted that departure time, day of week, and gender are indicators (dummy variables). The estimation results are summarized in Table 2. Variables with a t-value of ±1.96 or greater will significantly affect the risk preference level at the 95% confidence level. A positive value implies that the explanatory variable increases the bound values of risk preference with an increase in its magnitude. All of the explanatory variables included in the lower bound regression model, except gender, are statistically significant with plausible signs. A possible reason is that most of the investigated travellers are male, while only 10% of the travellers are female. In the upper bound regression model, all of the explanatory variables, except age and gender, are

statistically significant. This implies that age is more sensitive to the lower bound than to the upper bound, which is plausible because the tolerability for travel time uncertainty is usually reflected by the lower bound.

Table 2. Parameter estimation results of regression model

Variables	Description	Lower bound Coef.	t	Upper bound Coef.	t
OD distance	Straight distance between origin and destination	0.000152	19.03	0.000259	15.79
Age	Age of traveller (years)	0.00617	4.05	0.00308	1.61
Departure time	Departure time (off-peak hours = 0, peak hours = 1)	0.0814	2.38	0.2971	3.91
Day of week	Day of week (weekday = 0, weekend = 1)	-0.0865	-2.28	-0.2731	-2.71
Gender	Gender (male = 1, female = 0)	-0.0799	-1.17	0.01164	0.14
Constant		-0.471	-4.62	0.4372	3.41
sample size	3777				
R-squared		0.294		0.262	
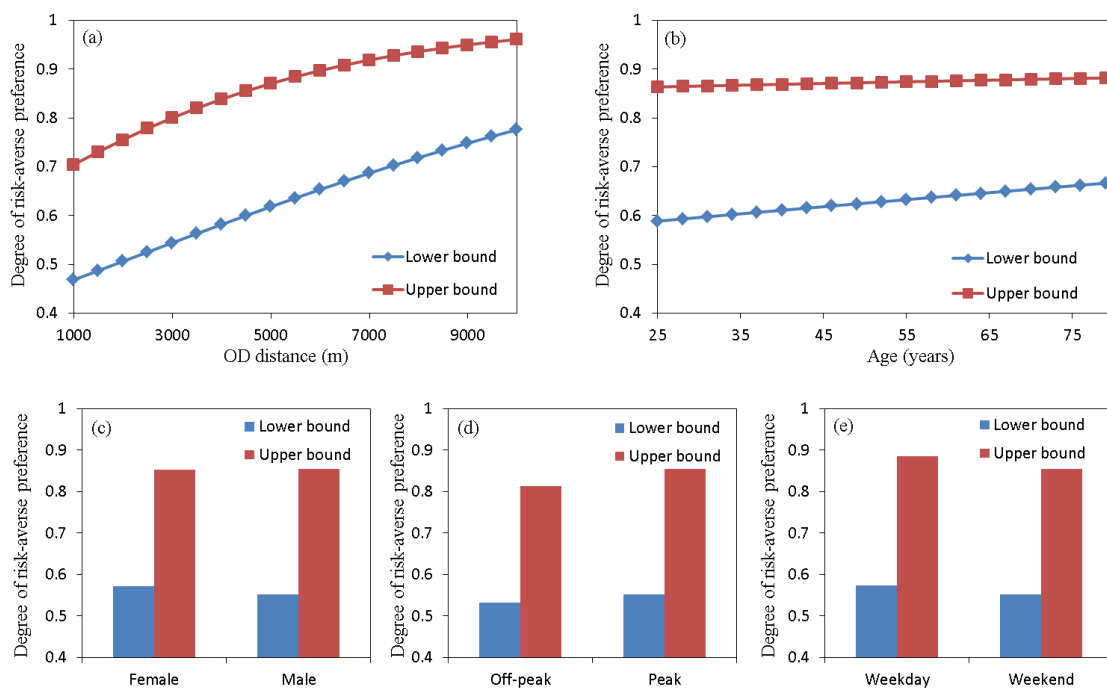
Figure 7. Sensitivity analysis results for each explanatory variable

The sensitivity analysis results shown in Figure 7 indicate how the bound values of risk preference change with a small change in the explanatory variables. The sensitivity analysis results for indicators such as gender, departure time, and day of week were calculated on the basis of the differences in risk preference when the indicator variable values were 0 and 1, with all other variables held constant at their mean values. The average OD distance was 5 km and the average age was 45 years in this dataset. The baselines for OD distance, age, gender, departure time, and day of week were set to 5 km, 45 years, male, peak hours, and weekday, respectively. As shown in Figure 7 (a), the sensitivity analysis indicates that as the OD distance increases, both the lower bound and the upper bound increase. This implies that travellers are more likely to prefer highly risk-averse paths with more reliable travel times. This is an interesting finding

because people usually prefer paths with more reliable travel times for long-distance trips, as the travel time variance will be more difficult to predict for such trips. The increase ratio is stable for the lower bound, while it decreases as the OD distance increases for the upper bound. More specifically, an increase of 1 km in the OD distance results in an average increase of 0.038 in the upper bound of risk-taking preference from 1 km to 5 km, while it only results in an average increase of 0.031 in the upper bound of risk-taking preference from 5 km to 10 km. As shown in Figure 7 (b), a unit increases in age differential results in a higher lower bound of risk preference, while such an effect is relatively small for the upper bound because age is not a significant factor in the regression model of the upper bound. The positive sign of age implies that as the age increases, people are more concerned about the travel time reliability when they plan a trip. The marginal effect of gender shown in Figure 7 (c) implies that female travellers have a higher likelihood of choosing a lower bound with a higher degree of risk preference, while male travellers have a higher likelihood of choosing an upper bound with a higher degree of risk preference. Because gender is not a significant factor in the regression models of the lower and upper bounds, we cannot conclude whether male or female travellers are more likely to choose a reliable path. As shown in Figure 7 (d), travellers departing during peak hours have a higher likelihood of higher risk-averse preference because both the lower bound and the upper bound increase compared to those during off-peak hours. This confirms that travellers prefer routes with highly reliable on-time arrival probability, especially when they commute to work. The marginal effect of day of week shown in Figure 7 (e) indicates that travellers are more likely to prefer highly risk-averse paths when they plan trips on weekdays. This is another interesting finding because a higher on-time arrival probability usually means a larger travel time budget. This implies that travellers are likely to choose a more reliable path even though they have to reserve a larger travel time budget when they depart on weekdays.

Although we have analysed the basic information of socio-demographic and trip characteristics using the regression model, some limitations have not been addressed. For example, the route choice behaviour might be influenced by other factors, such as travel purpose, driving comfort, distance, and familiarity. A sensitive survey on the preference of travel time reliability could confirm the validity of the proposed method. On the other hand, additional factors, such as trip purpose, income, habit, and weather, could be incorporated into the regression model.

4. Conclusions and future work

To search for an α-reliable path for risk-averse navigation in a stochastic network, it is necessary to provide the default value of the traveller's risk preference (α value). This paper introduced a new methodology for extracting the traveller's risk preference. Compared to questionnaire-based methods, the proposed method can be conveniently applied to traveller behaviour survey and preference analysis by using GPS trip records. The lower and upper bounds of the risk preference were estimated by stochastic dominance theory and α-reliable paths. Furthermore, a regression model based on a logistic function was employed to explore how individual properties (gender, age) and pre-trip information (OD distance, departure time, day of week) influence the risk preference. Sensitivity analysis of the regression model indicated that travellers are more likely to prefer highly risk-averse paths with more reliable travel times as the OD distance and age increase. Moreover, travellers departing during peak hours on weekdays are more likely to exhibit highly risk-averse preferences. However, gender was not found to be sensitive to the risk preference.

This study has some limitations. For example, only GPS data and basic personal information were used to estimate the travellers' risk preferences. We plan to conduct a questionnaire-based study to further validate this method in the future. Additional factors, such as traffic conditions and trip purpose, will also be incorporated to improve the accuracy of the proposed method.

References

Abdel-Aty, M. A., Kitamura, R., & Jovanis, P. P. (1995). Investigating effect of travel time variability on route choice using repeated-measurement stated preference data. Transportation Research Record, (1493), 39-45.

Ben-Elia, E., Di Pace, R., Bifulco, G. N., & Shiftan, Y. (2013). The impact of travel information's accuracy on route-choice. Transportation Research Part C: Emerging Technologies, 26, 146-159.

Chen, A., & Ji, Z. (2005). Path finding under uncertainty. Journal of Advanced Transportation, 39(1).

Chen, B. Y., Lam, W. H., Sumalee, A., Li, Q., Shao, H., & Fang, Z. (2013). Finding reliable shortest paths in road networks under uncertainty. Networks and Spatial Economics, 13(2), 123-148.

Chen, B. Y., Shi, C., Zhang, J., Lam, W. H., Li, Q., & Xiang, S. (2016). Most reliable path-finding algorithm for maximizing on-time arrival probability. Transportmetrica B: Transport Dynamics, 1-17.

Carrion, C., & Levinson, D. (2012). Value of travel time reliability: A review of current evidence. Transportation Research Part A: Policy and Practice, 46(4), 720-741.

Carrion, C., & Levinson, D. (2013). Valuation of travel time reliability from a GPS-based experimental design. Transportation Research Part C: Emerging Technologies, 35, 305-323.

de Palma, A., & Picard, N. (2005). Route choice decision under travel time uncertainty. Transportation Research Part A: Policy and Practice, 39(4), 295-324.

Lam, T. C., & Small, K. A. (2001). The value of time and reliability: measurement from a value pricing experiment. Transportation Research Part E: Logistics and Transportation Review, 37(2), 231-251.

Li, H., Guensler, R., & Ogle, J. (2005). Analysis of morning commute route choice patterns using global positioning system-based vehicle activity data. Transportation Research Record: Journal of the Transportation Research Board, (1926), 162-170.

Lima, A., Stanojevic, R., Papagiannaki, D., Rodriguez, P., & González, M. C. (2016). Understanding individual routing behaviour. Journal of The Royal Society Interface, 13(116), 20160021.

Miwa, T., Kiuchi, D., Yamamoto, T., & Morikawa, T. (2012). Development of map matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies, 22, 132-145.

Prato, C., Rasmussen, T., & Nielsen, O. (2014). Estimating value of congestion and of reliability from observation of route choice behavior of car drivers. Transportation Research Record: Journal of the Transportation Research Board, (2412), 20-27.

Papinski, D., Scott, D. M., & Doherty, S. T. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. Transportation Research Part F: Traffic Psychology and Behaviour, 12(4), 347-358.

Papinski, D., & Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. Journal of Transport Geography, 19(3), 434-442.

Sen, S., Pillai, R., Joshi, S., & Rathi, A. K. (2001). A mean-variance model for route guidance in advanced traveler information systems. Transportation Science, 35(1), 37-49.

Sivakumar, R. A., & Batta, R. (1994). The variance-constrained shortest path problem. Transportation Science, 28(4), 309-316.

Small, K. A., Winston, C., & Yan, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. Econometrica, 73(4), 1367-1382.

Wolf, J., Oliveira, M., & Thompson, M. (2003). Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey. Transportation Research Record: Journal of the Transportation Research Board, (1854), 189-198.

Xing, T., Zhou, X., 2011. Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach. Transportation Research Part B: Methodological, 45(10), 1660-1679.

Zeng, W., Miwa, T., Wakita, Y., & Morikawa, T. (2015). Application of Lagrangian relaxation approach to a-reliable path finding in stochastic networks with correlated link travel times. Transportation Research Part C: Emerging Technologies, 56, 309-334.

Zeng, W., Miwa, T., & Morikawa, T. (2016a). Prediction of vehicle CO 2 emission and its application to eco-routing navigation. Transportation Research Part C: Emerging Technologies, 68, 194-214.

Zeng, W., Miwa, T., & Morikawa, T. (2016b). Application of hyperpath strategy and driving experience to risk-averse navigation. IET Intelligent Transport Systems, 10(5), 338-346.