



Evaluation of Travel Time Reliability using "Revealed Preference" Data & Bayesian Posterior Analysis

Sida Jiang, Christer Persson and Karin Brundell-Freij

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 8, 2019

EVALUATION of TRAVEL TIME RELIABILITY USING “REVEALED PREFERENCE” DATA & BAYESIAN POSTERIOR ANALYSIS

Sida Jiang ^{a,1}, Christer Persson ^b, Karin Brundell-Freij ^c
^aWSP Sweden 121 88

Arenavägen 7 Stockholm-Globen, Sweden

¹ E-mail: sida.jiang@wsp.com, Phone: +46 (0) 70-231 68 77

^b Urban Planning & Environment, Royal Institute of Technology KTH
Teknikringen 10, Stockholm, Sweden

^c K2 – National Science Centre for Public Transport, Lund University
Bruksgatan 8, Lund, Sweden

Abstract

In Swedish context, the value of delay is deemed equalling to the value of travel time reliability (VoR), which is a factor of value of travel time (VoT) and mostly derived from Stated Preference (SP) studies. According to our knowledge, there are several issues with the SP method for obtaining VoR, for example, its deficiency in harmonizing the stated choices with the actual choices. On the other hand, Revealed Preference (RP) data from ticket sales has its limit in, for example, socioeconomic information of travellers and scenario variation.

This project aimed to use a RP method to evaluate travel time reliability through reliability ration (RR) - the relation between VoR and VoT upon several selected railway corridors, with Bayesian posterior analysis to infer socioeconomic differences between passengers given on their actual choices.

The data in the study are from two sources, ticket data from a major passenger operator SJ, and data on train movements from Trafikverket's (Swedish Transportation administration) TFÖR database. Both data sources are for the whole year 2009. The data includes 60 545 individual observations on traveler's route choice for two specific trip relations. The chosen trip relations are long-distance non-commuting trips with travel distances between 200 and 250 kilometers.

The project is a “proof-of-concept” for possible use of ticket data for the evaluation of travel time reliability. We can conclude that the estimated VoR – 1,13 times value of travel time, is in compliance with results from previous international studies using SP and/or RP data. The simulated distribution of RR from posterior analysis also clearly indicates a bimodal pattern of valuing travel reliability, probably due to socio-economic characteristics or trip purposes.

Keywords

Value of reliability (VoR), Bayesian model, posterior analysis, mixed logit, revealed preference (RP).

1. Background

With the deregulation of railway operations and modern information system practiced in an increasing manner in Sweden, travelers have more accessibility and flexibility yet inevitably more complicated travel choices to make from time to time. In this context, travelers encounter reliability issues in the form of delays when using railway. There are reasons to assume that travelers evaluate reliability in form of the expected day-to-day variability differently from delays relating to unexpected or surprising events. This study is an attempt to provide estimates of the valuation of travel time reliability, in economic sense. Areas of application for these valuations are in planning of maintenance measures and informing travelers about the rescheduling.

According to the latest socio-economic evaluation guide from Swedish National Transport Administration (Trafikverket, ASEK6.1), value of reliability (VoR) is 3.5 times VoT which has been derived from stated preference (SP) studies. Nevertheless, there are issues with the SP method for obtaining VoR. For example, travel reliability that is a measure of a probability distribution is not a straight forward concept to present to the respondents; also, it remains unclear whether there is a robust consistency between choice that are made in a specific hypothetical situation and revealed behavior observed in reality. These issues would be even larger for this proposed study where we need to distinguish between expected reliability and unexpected delays, and the former will be modelled in as a perceived factor determining how traveler choose ahead of different alternatives, while the latter can't be counted before the trip is made.

Due to the above-mentioned issues with the SP method, we have chosen to use revealed preference data recorded in statistics from year 2009. The data in the study are from two sources, ticket data from a Swedish railway operator (SJ) and data on train movements from Trafikverket's TFÖR database. The main reason for using relatively old data is that it has given us the opportunity to use detailed ticket data without further concerns on commercial disclosure. The data includes 60 545 individual trips on traveler's route choice for two specific trip relations: 1) from Örebro to Stockholm and 2) from Borlänge to Stockholm. These two relations have been chosen to homogenize the data. Homogenization of data is a reasonable strategy for a new research area since it will make estimated valuations more accurate, but in the same time compromises the generalizability of these estimates to trip relations with other characteristics than the relations contained in the sample. The chosen trip relations can be described as long-distance non-commuting trips with travel distances between 200 and 250 kilometers.

2. Hypothesis and Model Design

2.1 Empirical Setup

To evaluate travel time and its reliability, we use revealed route preference for railway traffic. This is a new area of research and it may be difficult to empirically establish the proposed valuation of travel time reliability from other determining factors. Therefore, the evaluation is based on a homogenized route choice data set consisting of three choice relations with the central station in Stockholm as one end-point. All three choice relations have similar trip distances 200-250 kilometers. The choice relations are depicted in figure 1 below.

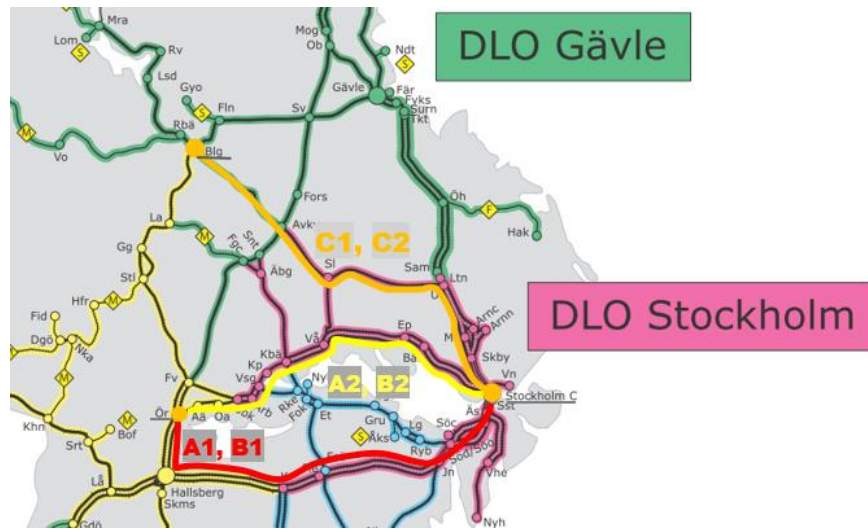


Figure 1 Illustration of studied railway routes

- A. From Örebro (Ör) to Stockholm Center (Stockholm C) in the morning peak hours: alternative 1 is a transfer via Hallsberg and alternative 2 is a non-transfer train,
- B. From Örebro to Stockholm C in the off-peak afternoon: alternative 1 is a transfer via Hallsberg and alternative 2 is a non-transfer train,
- C. From Borlänge (Blg) to Stockholm C in the morning peak hours: alternative 1 is regional train and alternative 2 is high-speed x2000.

For instance, in choice task A, travelers can only choose either alternative 1 or 2, and both alternatives can be characterized by planned travel time PTT, departure delay (usually informed before departure from the start station) D^{inf} and the uninformed but self-estimated travel time reliability – standard deviation of travel time $SD(TT)$, travel cost C and alternative specific constants ASC . A linear form is assumed for the specification of the utility for a traveler, which therefore attain the following form:

$$U_1 = \beta_T * PTT_1 + \beta_{SD} * SD(TT_1) + \beta_{D^{inf}} * D_1^{inf} + \beta_C * C_1 \quad (1)$$

$$U_2 = \beta_T * PTT_2 + \beta_{SD} * SD(TT_2) + \beta_{D^{inf}} * D_2^{inf} + \beta_C * C_2 + ASC \quad (2)$$

Travel time PTT is the planned travel time or interchangeably timetable travel time. PTT is rather fixed and can be considered as only varying over alternatives and choice tasks, thus the difference between PTT over two alternatives can function as choice-task specific constants (CTSC), and due to its importance in explaining the utility and to prevent high correlation with ASC, PTT (or CTSC) is employed to replace ASC.

Travel cost C is not available at the individual purchase level therefore been assumed the same and quantified with an average over all individual tickets in the studied alternative, and in this case it is missing thus enters into the constant term ASC.

Travel time reliability or the risk of delay against time table is difficult to model and results in a massive variety of indicators for theoretical and practical appliance. Börjesson & Eliasson (2011) concluded evaluation of travel time reliability varies over the frequency of travel delay; Fosgerau and Karlström (2010) are giving an expression for VoR when standard deviation is used as the attribute for reliability. The expression for VoR, in this framework, is a function of the travel time distribution. In accordance with (Fosgerau and Karlström, 2010) this study uses standard deviation of travel time as the indicator of travel time reliability. However, in contrast to their approach, (i. e. to estimate the travel time distribution and then compute VoR from the distribution), we estimate reliability ratio RR empirically, as the ratio β_{SD}/β_T between the parameters given by the utilities in equation (1) and (2). Since standard deviation is by far the most common attribute for travel time variability, which makes the results comparable with other SP/RP studies in Sweden and worldwide. Yet it is not informed or deterministic, which invites random effects from person to person, together with the needs of posterior analysis, the project therefore employs also mixed logit model for better design and modal fit.

2.2 Mixed Logit Model

The utility equation for mixed logit (McFadden & Train, 2000) with randomized effects of travel time and travel time reliability assumed to follow a normal distribution over the travelers, so that the unconditional choice probability:

$$P_{n, i}(\Omega) = \int_{\beta} \left[\frac{e^{\beta x_i}}{\sum_{j=1}^J e^{\beta x_j}} f(\beta_T | \Omega) \right] d\beta \quad (3)$$

where $f(\cdot | \Omega)$ is the density of a normal distribution with parameters Ω . The difference from ground model, a binomial logit with utilities given by (1) and (2), is that the choice probabilities $P_{n, i}(\Omega)$ is the average binomial choice probabilities over the normal distribution with parameters Ω . To establish a computationally efficient estimation of this mixed model, uniform random draws are first generated using a Halton sequences¹, then the inverse function of the cumulative density function is used to derive standardized **normal draws** with mean 0 and standard deviation 1. With this standardized normal distribution, the utility functions with randomized effects of travel time and travel time reliability can be written as follows:

$$U_1 = (\beta_T + \sigma_T * \mathbf{draws}_T) * PTT_1 + (\beta_{SD} + \sigma_{SD} * \mathbf{draws}_{SD}) * SD(TT_1) + \beta_{D^{inf}} * D_1^{inf} \quad (4)$$

$$U_2 = (\beta_T + \sigma_T * \mathbf{draws}_T) * PTT_2 + (\beta_{SD} + \sigma_{SD} * \mathbf{draws}_{SD}) * SD(TT_2) + \beta_{D^{inf}} * D_2^{inf} \quad (5)$$

That the choice probability can therefore be reformulated with $N(0,1)$ random draws as follows:

¹ Halton method divides 0 – 1 space into p_k segments (with p_k giving prime used as base for parameter k), and by systematically filling the empty spaces, using cycles of length p_k .

$$P_{n, i}(\beta_x, \sigma_x) = \int_{\varepsilon_x} \left[\frac{e^{\beta x_i + \sigma_x \varepsilon_x x_i}}{\sum_{j=1}^J e^{\beta x_j + \sigma_x \varepsilon_x x_j}} \phi(\mathbf{0}, \mathbf{1}) \right] d\varepsilon_x \quad (6)$$

Instead of estimating fixed parameter for travel time and travel time reliability, both the mean β_T , β_{SD} and standard deviation σ_T , σ_{SD} are to be estimated based upon simulated log-likelihood maximization (Bhat, 2001).

3. Model Estimation

The travel time and travel time reliability has been randomized with standard normal draws, according to the preceding sections, so that the mixed logit model requires estimation of both mean and sigma for how the sensitivity of travel time and travel time reliability vary over individuals. Also, the mixed logit model is a non-linear model which can result in numerous local optima. To handle the local optima issue, different initial values has been run for the first 100 iterations and the initial values with the best log-likelihood has been chosen for further model estimation. The initial log-likelihood is LL(0): -41966.6 and the model converged at **LL(final): -24 151.5**.

Table 1 Estimation of Mixed Logit Model (with 300 random draws)

	<i>Est.</i>	<i>s.e.</i>	<i>t-value(0)</i>	<i>robust s.e.</i>	<i>robust t-value(0)</i>
β_{SD}	-0.46	0.01	-48.98	0.01	-46.72
β_T	-0.37	0.01	-49.59	0.01	-46.97
σ_{SD}	-0.25	0.01	-29.10	0.01	-23.64
σ_T	-0.31	0.01	-40.95	0.01	-33.44
$\beta_{D^{inf}}$	0.17	0.00	48.22	0.01	21.68

Rho-sq: 0.42; adj. rho-sq: 0.42; AIC: 48 312.9

Standard deviation σ for both travel time (T) and travel time reliability (SD) are (strongly) significant different from 0. This means that we have shown that there is variation across individuals of how those attributes as trade-off against each other. Including a random effect can hence improve the modal fit. The same conclusion can be found by comparing LL(final), adjusted rho square and AIC that mixed logit model fits better than corresponding basic logit model.

Coefficient for departure delay (D^{inf}) is positive, which is contrary with the expectations. the likely explanation is that the explanatory power of departure delay is likely to be confounded by other variables that varies over choice tasks. Because of this problem, which is inherent in data, the project conclusions focus more on the importance of planned travel time and travel time reliability. Again, this counter-intuitive finding is most probably limited to the specific study scope and in the many cases with only small departure delay. In the rarer cases where there is a long departure delay, it is most probable that departure delay will be used (righteously) as a predictor by the traveler, indicating that one can expect

a longer than usual delay at the final destination also.

4. Posterior Analysis

The mixing distributions given by the parameters in table 1 can be used directly to construct the distribution of the reliability ratio RR. However, in the specification, the mixing distributions are assumed to be independent. This assumption is likely to not be fully valid. Since the distribution of especially very high RR as well as very low RR among individuals is highly dependent on the validity of this assumption, the computed RR from the mixing distributions can be expected to deviate considerably from its true value in the population. Further the mixing distributions are assumed to be Gaussian. Under the estimated parameters given in table 1 there will, for example, be a sizeable proportion of the population for which RR will have a wrong sign. Therefore, it is a need for a more robust estimation of RR. The method used is a so called posterior analysis (Hess, 2007) where, in a Bayesian spirit, the posterior distribution of RR is obtained by applying the mixing distributions to the likelihood of the data. This method can be seen as a way to correct the mixing distributions such that they comply to observed dependencies in the data (i.e. the likelihood). In this sense, the obtained distribution for RR can be seen as more robust than the distribution obtained directly from the estimation of mixed logit model.

Knowing that different individual has significantly different evaluation of both travel time and travel time reliability, we can further divide the individuals into several groups conditional on their observed choices. In the meantime, the reliability ratio is no longer limited to the average level as earlier illustrated ratio of coefficients, by using posterior analysis upon the mixed logit model. Each individual is assumed to follow a random distribution (with simulated random draws) and each individual is assigned with the expected values of this random distribution termed as conditional mean. Notice that conditional mean is not the actual sensitivities for that individual but the expected mean, in other words, it is associated with different simulation of corresponding distribution and how many random draws one allows for each individual. The study uses 300 as number of random draws, furthermore sensitivity to both travel time and travel time reliability is assumed to be normally distributed.

For more details in posterior analysis of mixed logit model please refer to Hess (2007). The probability of observing the specific value of β given the choice of individual n :

$$\widehat{\beta}_n = \frac{\sum_{r=1}^R [L(Y_n|\beta_r)\beta_r]}{\sum_{r=1}^R L(Y_n|\beta_r)} \quad (7)$$

Where β_r with $r = 1, \dots, R$ are i.i.d draws, this method will relax the independent assumption of composing variables imposed by unconditional estimation, in other words, the resulting ratio of estimated coefficients is supposed to be more robust and fit into the reality revealed by the data. However, this would again lead to problems with data outliers.

The descriptive statistics of posterior analysis results for travel time, travel time reliability and reliability ratio is summarized in table 2:

Table 2 descriptive statistics of posterior analysis results for the parameters for travel time, travel time reliability and reliability ratio

<i>Statistics</i>	<i>Travel time</i>	<i>Travel time reliability</i>	<i>Reliability ratio</i>
1 st quartile	-0,207	-0,252	0,806
median	-0,106	-0,213	1,280
mean	-0,101	-0,214	1,132
3 rd quartile	-0,040	-0,197	3,808
Std. dev. of Sample	0,137	0,050	94,517
Std. dev. of Sample mean	0,0006	0,0002	0,384

As the table shows, the mean and median of RR from mixed logit model are both slightly above 1 and quite close to RR value from the recent international studies summarized by Carrion and Levinson (2012), see the figure 2 bellows.

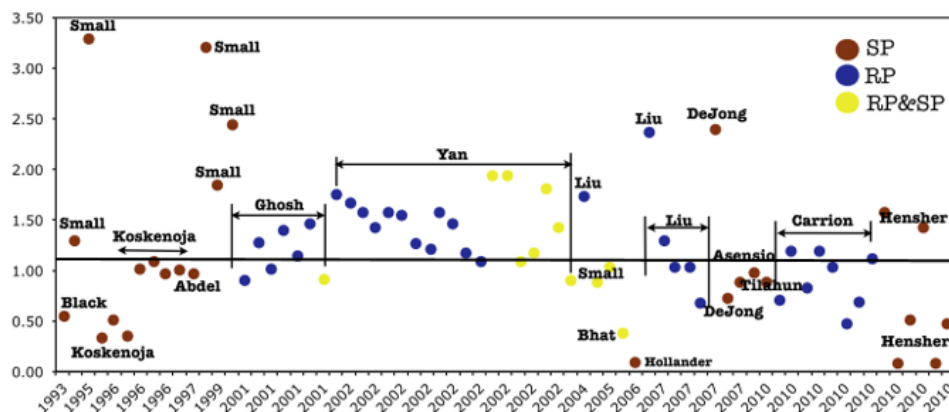


Figure 2 reliability ratio of selected studies (Carrion and Levinson, 2012)

One may argue, based on the figure, that estimated reliability ratios have declined, over the past two decades, in both SP and RP data, with RP estimates more constrained in middle of the span. In other words, travel time reliability seems to play relatively less and less role compared to travel time in the utility function. This may be explained by easier accessibility to travel information in general: travelers can forecast the coming journey and make changes accordingly, so that unreliable travel time become more and more predictable; also, substitution with digital activity can help travelers make as efficient use of travel delay as she/he can make of planned travel time. This declining trend of reliability ratio should probably be considered when conducting cost-benefit analysis (CBA) with travel time reliability involved. Results from our model (illustrated as a line in figure 2) is quite in line with other RP studies in which RR has varied from 0.5 to nearly 2.5.

But also, from posterior analysis, a significant spreading of RR between 1 and 4 has been observed, which could potentially and partly be explained by different trip purposes.

This variation needs to be further examined and put in relation to complementary information about trip. The detailed distribution of the estimated reliability ratio is indicated in the figure 3:

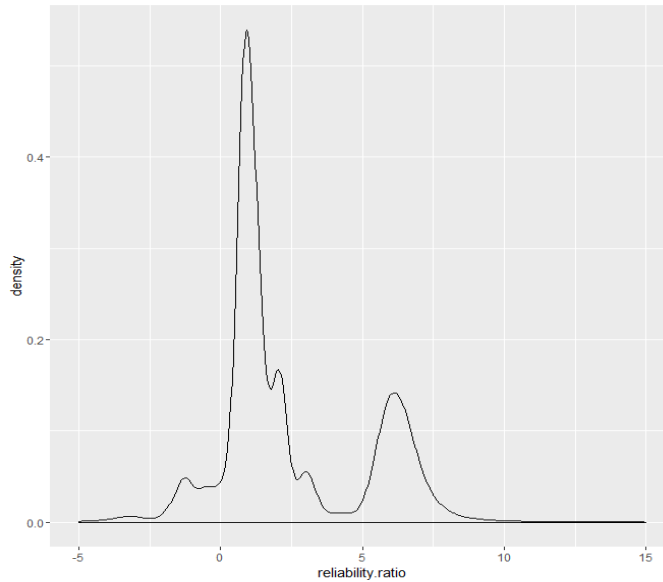


Figure 3 Probability density function of RR from posterior analysis

As expected, the major part of distribution is positive with several modes, the 1st group with around 50% of the individuals has a RR slightly less than 1. This group of travelers are the ones for which their observed choices indicate that they are relatively less sensitive towards travel time reliability. This group is also the majority in our data, but different data and scope of study can change its dominance and thus yields very different statistics. One important point of our results is therefore that multiple clusters or groups can be seen clearly in the posterior analysis, and we may need further data such as SP to understand better how socio-economic variables or trip purpose divide the sample, and then to specify reliability ratio with respect to e.g. private/business/work trip.

The 2nd largest group with around 15% of individuals has a high RR, with an average very close to 6. This group is thus about 6 times more sensitive of travel time reliability than the overall average. The division into two groups is however not absolute: the analysis also suggests that there are also individuals with RR between 1 and 6.

As with all other results, our conclusions are limited to the range of travel distances for which we have data. Obviously, the magnitude of RR can vary with among others travel distance, and the results illustrated above can only draw insights about the trips with distance between 200 and 300 km. Nonlinearity of RR with respect to travel distance can be complemented with data of other routes at different length.

5. Conclusions

In this project, ticket sales data from SJ has been transformed and treated as revealed preference data combined with travel time, travel time reliability and departure delay from

TFÖR database. Both data sources can be obtained from historical database and studies over different routes and years can be conducted for other analysis purposes, but one can also – as was done in this study – use the data as “RP” (revealed preferences) to conceptually examine how travelers react to travel time uncertainty or different forms of delay. Main results of this project can therefore be argued to be that it proved possible to use the type of data on observed behavior (RP-data), to estimate a model of how uninformed delays (travel time uncertainty) affect individual travel behavior.

As a basis for our analyses, we have used the choice task when a traveler chooses between different scheduled travel options. (One of the drawbacks of this approach is that the behavioral response not to go by train is not included). Our analysis circles around the travelers’ trade-off between three central qualities of the travel options: Travel time (as planned in the time table), delay at departure, travel time uncertainty (based on the distribution of real travel times in the recent past)

Our data only comprises three distinct choice situations (Choice situation = A pair of adjacent scheduled train departures for the same destination). Although the choice tasks have been selected so that they are similar in nature, the alternatives will inevitably differ in many more aspects than the three measured qualities that is introduced in our analyses. Therefore, there is a risk that our results are confounded by other variables with which our explanatory variables co-vary over alternatives and choice tasks. Also, only travel distances in the range 200-300 km is covered in data.

For travel time (as scheduled in time table) our estimations give the intuitively correct sign for the estimated parameter. Travel time varies only with single minutes for the same alternative in a given choice situation (due to minor modifications of the timetable during the observed year). Thus, the estimated value is based almost entirely on the variation between the three choice tasks. Therefore, it is reassuring that the estimated value has the correct sign. Never the less, it is clear that a larger data set (that is many more choice tasks) would have been highly desirable. Since it was not possible to estimate parameters for the sensitivity to costs, it is not possible to check whether the estimated sensitivity to travel time is reasonable in terms of “value-of-time”. However, we can conclude that the ratio between the parameters estimated for travel time and travel time reliability, was estimated to 1.13, which is very much in line with what has been estimated in previous studies. The reliability ratio can be used for socio-economic evaluation regarding investment and maintenance for a more robust railway service, from the perspectives of passengers.

6. Future Work

Before our study was conducted, the novel approach we were proposing raised concerns as to whether:

1. Data quality was good enough, given that data from multiple sources were combined and one data source (ticket statistics) had not previously been used.
2. It would be possible to estimate reliable parameter values for the two relevant variables journey time and travel time uncertainty.

We can now conclude that data quality seems to be sufficient, and that the data allow the estimation of models that are suitable for the purpose.

A particular difficulty is that we have studied how travelers choose between trains. This means that we miss in our analysis the traveler’s option to adapt to uncertainties and delays by abandoning the train altogether, either by switching to another mode, or to forgo the trip.

If future studies are extended to include also such alternatives, it may help to allow the estimation of more relevant parameters.

To summarize what is probably needed to increase the possibility to estimate the effects / parameters for informed delay also:

- Allow the option "not-use-train" into the described choice situation. In practice, this can be done by using not only distribution of rail passengers between train alternatives, but also the total number of train travelers.
- Include more data (more choice situations) to provide (1) better estimates, (2) reduced risk of unmeasured attributes that vary between choice tasks being confounded into the estimated parameters and (3) possibility to study how values differ between different types of travel (there are indications of the estimates on the valuation of uncertainty is bimodal).

Mixed logit model has been tested with better fit for the data. In future work it would also be useful to develop that approach further, for example test different random distribution, number of draws as well as modified specifications of utility function for improvement of model fit. In a word, current results have shown differences between how travel time, travel time uncertainty and different types of delay is evaluated, and also significant variance cross observed individuals. Future analysis is to extend the model so that it can utilize RP data to calculate VoT, VoR and RR over different trip purposes, travel distance and for other analysis practices.

References

- Bhat C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model [J]. *Transportation Research Part B: Methodological*, 35(7): 677-693.
- Börjesson M., Eliasson J., 2011. On the use of "average delay" as a measure of train reliability [J]. *Transportation Research Part A: Policy and Practice*, 45(3): 171-184.
- Carrion C, Levinson D, 2012. Value of travel time reliability: A review of current evidence [J]. *Transportation research part A: policy and practice*, 46(4): 720-741.
- Fosgerau, M., Karlström, A, 2010. The value of reliability, *Transportation Research Part B: Methodological* 44, no. 1, 38-49.
- Greene W.H., Hensher D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit [J]. *Transportation Research Part B: Methodological*, 37(8): 681-698.
- Hess S., 2007. Posterior analysis of random taste coefficients in air travel behaviour modelling [J]. *Journal of Air Transport Management*, 13(4): 203-212.
- McFadden, D., & Train, K. 2000. Mixed MNL models for discrete response. *Journal of applied Econometrics*, 15(5), 447-470.
- Trafikverket, 2018. Analysis method and socioeconomic calculation values for transport sector: ASEK 6.1. Chapter 8 cost for congestion and delay. Table 8.1.
- Walker J., Ben-Akiva M., 2002. Generalized random utility model [J]. *Mathematical Social Sciences*, 43(3): 303-343.