A meta-analysis of travel time reliability

Yin-Yen Tseng*

Department of Spatial Economics, Vrije Universiteit Amsterdam

Abstract

The reliability and scheduling delay of travel time attributes have been considered as important factors in traveler's decision making. Numerous studies have attempted to incorporate travel time reliability and scheduling delay early/late attributes into traveler's choice models since the last decade. However, there is still a wide-ranging debate on empirical valuations, and substantial differences of estimation values are shown among studies. Our aim in this study is to investigate several unresolved issues in the empirical valuation of reliability and scheduling delay delay/late and estimate these effects by means of a multivariate statistical technique: meat-analysis. The main finding is that including all reliability and scheduling delay early/late attributes in choice model would lead to lower estimated values for these attributes. We also find that the stated preference data produce substantial lower values for the ratio between scheduling delay early/late and travel time coefficients and the possible explanation may be the misperception error together with the risk aversion attitude of travelers.

Keywords: travel time reliability, scheduling delay early, scheduling delay late, meta-analysis.

1 Introduction

Various factors are known to govern travel behavior. Along with the attribute of travel time, reliability[1] has been regarded as an important component in individual's decision making of route choice or mode choice. Intuitively, the concept of 'reliability' suggests that an individual has to make his or her travel decision under uncertain circumstances with respect to travel time, and hence the nature of reliability can be described by the distribution of travel time (Bates, 2001).

A fair number of studies have attempted to incorporate travel time reliability attributes into travelers' choice models during the last decade. However, there is still a wide-ranging debate on reliability valuation, particularly in the way of modeling; and substantial differences of estimation values are shown among studies. No consensus has been achieved thus far, neither on point estimates nor on the methodological question of how to measure the value of reliability.

*Tel: +31-20-444-6098; Email address: ytseng@feweb.vu.nl

[1] Several transport attributes can also be referred to reliability, such as reliability of level of service of road, reliability of transport facility, and reliability of traffic congestion. In this study we restrict our attention to reliability on travel times.

In this study we focus on the review of empirical estimates of reliability in travel time related attributes. We look not only at the valuation of travel time reliability itself, but also concern the valuation of scheduling delay variables. Our aim is to study the sources of variations in empirical estimates and to investigate the unresolved issues by means of meta-analysis, a quantitative method of literature surveys. By performing the meta-regression, the main difference in estimates can be explained in a systematic way. Thus, the merit of meta-analysis may serve as the guideline for future research in this area.

The paper is organized as follows. Section 2 considers the concepts of value of time, reliability, and scheduling cost. It also shows the most used empirical modeling approach in travel time reliability valuation. Section 3 discusses the main arguments and possible sources of variation in empirical works. Section 4 describes the data and shows the overview of empirical estimates in the context of various reliability indications such as the reliability ratio, scheduling delay early ratio and scheduling delay late ratio. The meta-regression results and discussions are presented in Section 5, and Section 6 concludes.

2 Theoretical framework

2.1 Empirical model

The conventional approach of modeling travelers' choice behavior is discrete choice analysis, which stems from utility maximization theory and assumes that respondents will select the alternative in the choice set that has the highest utility. Among various models used in discrete choice analysis, the *random utility* (RU) model is the most intensively used one in empirical assessment of travel behavior. In such an approach, the utility of individual *i* from choosing alternative *j* is given (Ben-Akiva and Lerman, 1997) by:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{1}$$

The first part V_{ij} of Eq (1) is the 'deterministic part' or 'systematic part' and is constituted by observed attributes of the alternative and characteristics of the individual, that is,

$$V_{ij} = f(X_{ij}, \beta_i) \tag{2}$$

where X_{ij} is the vector of attributes as perceived by individual *i* for alternative *j* and β_i reflects the characteristics of individual *i*.

The choice of functional form of *f* is very general. The most basic model is the linear additive form, represented as,

$$V_{ij} = \sum_k \beta_{ik} x_{ijk} \tag{3}$$

where subscript *k* represents the set of attributes that may affect individual's utility in choosing alternatives *j*.

The second part ε_{ij} of equation (1) is the random (or error) term, which is unobserved by the researcher. Various models can be derived from different assumptions as the error term distribution. In practice, the most popular one is the logit family, which assumes the error term follows extreme value type 1 distribution. The advantage of logit model is its tractability, though it imposes restrictions on the covariance structure of error terms. Thus, many models deviating from standard logit, such as nested logit and generalized extreme value logit, have been developed, and aim to relax the restrictions on error terms.

2.2 Concepts of reliability and scheduling delay

Since the concept of reliability can be regarded as the distribution of travel time, it appears that at least two dimensions of travel time have to be considered in modeling the effect of reliability—namely, its magnitude and frequency. One plausible indicator of reliability is the variance or standard deviation of travel time, which can be evaluated in practice to illustrate the loss of utility due to the amount of this value.

Along with the utility loss incurred by the unreliability in travel time, a traveler may also attach additional (dis-)utility to arriving at the destination before or after his preferred arrival time (PAT). Thus, the difference between actual arrival time and preferred arrival time may play a role in traveler's decision making. Following Small's paper (1982), this measurement of difference between PAT and actual arrival time is defined as schedule delay (SD). That is, $SD = PAT - [t_h + T(t_h)]$, where t_h is the departure time and the amount of travel time $T(t_h)$ depends on the chosen departure time. Fig.1 shows the relations between departure time t_h, travel time $T(t_h)$ and preferred arrival time (PAT).

In general, people may value early and late arrivals differently according to the different consequences. Most research (Small 1982, Noland and Small 1995, Bates et al. 2001) evaluate SD as two separate terms, schedule delay early (SDE) and schedule delay late (SDL), which can be expressed as:

$SDE = Max(0, PAT - [t_h + T(t_h)])$ and $SDL = Max(0, [t_h + T(t_h)] - PAT)$

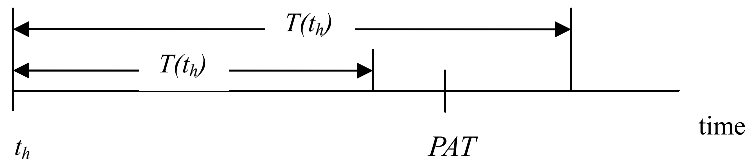

Figure 1 Departure time, travel time and preferred arrival time demonstration

2.3 Modeling approaches

The earliest work to consider the effect of reliability in travel behavior originates from the mean-variance approach. Jackson and Jucker (1981) specified a model where a traveler can make the trade off between travel time and variance of travel time explicitly. Both of these two elements are included in a cost function that travelers seek to minimize it. A general form of this mean-variance approach is given by Eq (4).

$$Min\ C = E(T) + \lambda \cdot Var(T) \quad (4)$$

The coefficient of variance of travel time λ can be seen as a measure for risk aversion. Instead of $Var(T)$ in Eq.(4), sometimes the standard deviation was used. The survey made by Jackson & Jucker contained a sequence of paired comparison questions, in which the respondents were asked to choose their preferred alternative. The results were shown as a distribution of the risk aversion coefficients λ. This mean-variance approach, used effectively in the field of portfolio analysis in financial market, has its sounded theoretical backgrounds and can be applied easily in mode or route choice. Yet the weakness of this approach might be its disability in dealing with departure time choice behavior with scheduling constraints, which will be discussed in the next paragraph.

Proposed by Small (1982), the scheduling concept was first modeled in traveler's choice behavior and examined with empirical data. The general form of indirect utility function can be presented as,

$$U = \alpha \cdot T + \beta \cdot SDE + \gamma \cdot SDL + \theta \cdot D_L \quad (5)$$

To introduce the concept of uncertainty, Noland and Small (1995) extend the scheduling model from Eq.(5) by considering the probability distribution of travel time and adding an additional random component. The result is presented as Eq.(6)[2]. This choice problem under uncertainty is what is called *Maximum Expected Utility* (MEU) theory.

$$E(U) = \alpha \cdot E(T) + \beta \cdot E(SDE) + \gamma \cdot E(SDL) + \theta \cdot P_L + \delta \cdot f(std) \tag{6}$$

The basic idea of Eq.(6) is that travel time reliability, regarded as a function of the standard deviation of travel time, may produce inconvenience in planning activities. Its effect should however, be made independent of scheduling concerns in the model. Note that the specification in Eq.(6) implies consideration of both the scheduling model and the mean-variance approach. A number of previous researches have attempted to examine the utility function derived from Eq. (6), and usually the cost term is also included in this model. Our main interest of analysis in this present paper will be the parameters of reliability, schedule delay early, and schedule delay late, all compared to the parameters of the travel time or cost term.

Once the model is estimated, one can find the marginal rate of substitution between any pair of the attributes in the bundle. The monetary value of travel time (VOT) is defined as the marginal substitution rate between travel time and costs and hence as the ratio of the respective coefficients. Similarly, the monetary value of reliability (VOR), value of schedule delay early (VSDE) and value of schedule delay late (VSDL) can be expressed as following (see Eq (6).)

$$VOT_i = \frac{\partial U_{ij} / \partial T_{ij}}{\partial U_{ij} / \partial C_{ij}} = \frac{\beta_T}{\beta_c}, \qquad VOR_i = \frac{\partial U_{ij} / \partial R_{ij}}{\partial U_{ij} / \partial C_{ij}} = \frac{\beta_R}{\beta_c},$$

$$VSDE_i = \frac{\partial U_{ij} / \partial SDE_{ij}}{\partial U_{ij} / \partial C_{ij}} = \frac{\beta_{SDE}}{\beta_c}, \qquad VSDL_i = \frac{\partial U_{ij} / \partial SDL_{ij}}{\partial U_{ij} / \partial C_{ij}} = \frac{\beta_{SDL}}{\beta_c} \tag{7}$$

where β_T, β_c, β_R, β_{SDE}, and β_{SDL} are referred to the coefficients of travel time, travel cost, reliability, schedule delay early, and schedule delay late variables in the estimated model respectively.

One practical issue in the meta-analysis that will follow is that some studies do not include the cost related terms in their estimated model. Yet these studies did include the variables we are interested in like travel time, reliability, SDE or SDL. To increase the number of observations in the database, we therefore decided to use the marginal rate of substitution between time and reliability for our variable of interest in meta-analysis. The

[2] See Noland and Small (1995) for the modeling details.

marginal rate of substitution between travel time and reliability is the so-called *reliability ratio*, i.e., $RR = \beta_R / \beta_T$, defined by Black and Towriss (1993). To facilitate the empirical analysis of scheduling variables, we also define *schedule delay early ratio* and *schedule delay late ratio* as $SDER = \beta_{SDE} / \beta_T$ and $SDLR = \beta_{SDE} / \beta_T$. Another advantage of using the reliability or scheduling ratio in the analysis is to get rid of the transformation problem because of the exchange rate conversion problems. Since the monetary values of reliability and scheduling variables are estimated based on the local currencies, and hence cannot be comparable with the original values. Using the free of unit ratio values will not be affected by the conversion procedure.

3 Issues in the valuation of reliability

3.1 Revealed versus stated preference

There are two major sources of preference data, revealed and stated preference, which both can be used to estimate discrete choice models. Traditionally, empirical studies of traveler's choice behavior rely on data from observing what people *actually* do, i.e., revealed preference (RP) data. However, recent studies favor data from people's choice under hypothetical situations, which we refer to as sated preference (SP) data. As Louviere et al. (2001) summarized, there are some compelling reasons for economists to use SP data; e.g., to estimate demand for new products with new attributes or features, to have sufficient variation of explanatory variable to allow for reliable model estimation, or when observed explanatory variables are highly collinear in the marketplace. The last one, above all, is the most common limitation in RP data that may cause severe identification problems in econometric analysis.

The most serious critique for SP data is probably its lack of reality, and doubts on the validity of hypothetical choices. Thus, SP data depends largely on the "contextual effect". However, many researchers believe that this problem can be solved by well-designed SP surveys. Indeed, well-designed SP data may be superior to ill-conditioned RP data, which are problematic in model estimation. In any case, our interest in the present study is not to argue which method is most appropriate to use, but instead to see whether there is systematic difference in the estimates between SP and RP data.

As we discussed above, the inherent difference of *RP* and *SP* type data may lead to some perception problem for respondents and also the estimation in econometric model. Earlier studies (Ghosh, 2001; Yan, 2002) show that the median SP estimates of VOR and

VOR are about half of the median RP estimates and the differences are statistically significant; while in some other fields, for instance, in environmental economics, the estimates from SP are expected to be higher than RP estimates (Lanoie et al., 1995). Brownstone and Small (2003) mention that the difference between SP and RP is probably caused by the misperception of travel time in RP survey, and people may exaggerate the amount delay time due to the impatience with heavy traffic. Whether the SP method underestimates our targeted estimates in a systematic way will be left for the meta-analysis.

3.2 Utility specification: reliability versus scheduling variables

UK studies, Arup 2002 and Bates et al. 2003, concluded that the value of reliability can be entirely explained by expected scheduling cost. Indeed, some empirical works (Noland, 1995; Small et al. 1999) obtained insignificant results in valuating the effect of reliability when including both of reliability and scheduling variables in the model. One plausible explanation could be that most empirical work does not distinguish between reliability and scheduling concepts very well in the context of questionnaire, and hence the respondents might mix up these two effects into one. Thus, the estimated scheduling costs usually also reflect the unreliability costs.

Though the concept of reliability and scheduling are closely related to each other, they should not be treated as identical. The idea is that apart from people's scheduling preference, they may have some additional disutility due to the inconvenience or anxiety caused by unreliability of travel time, even when the 'expected scheduling delay' cost is the same. Moreover, a great part of trips do not have strict scheduling constraints (e.g., shopping or leisure) and people may be indifferent as long as arriving at the destination within a certain range of arrival times. In such a case, the disutility may come from the inconvenience of planning due to the unreliability of travel time rather than scheduling considerations.

Another subtle but relevant point in utility specification issue is the inclusion of lateness variable, P_L in Eq. (6), which can be modeled as either the probability or the dummy of lateness. Similar to the argument of reliability versus scheduling variables, we can infer that the existence of lateness variable in the model may probably affect the estimates of reliability and scheduling variables in the same manner, and particularly for the scheduling delay late variable due to the closely relation between P_L and *SDL*.

One of our main purposes is to investigate this utility specification effect and to see what the extent of this influence is.

3.3 Types of choice set

It remains unclear whether the estimations of reliability or scheduling variables would be varied in different types of choice set. (e.g., route choice, mode choice, or the combination of departure time choice). Briefly speaking, the characteristics of choice problems are distinct in some points between 'within' mode choice (i.e., route choice) and 'between' mode choice (i.e., mode choice); whereas the departure time choice can be incorporated into any of these two type of choice models by explicitly indicating the departure and arrive time in the choice questions. Thus, the utility setup--the set of attributes included in the model--should be able to respond to these different features of choice problems.

Basically, if the underlying utility function is correctly specified to reveal traveler's actual choice behavior, the estimates of reliability and scheduling variables obtained from different choice set domains should be close to each other. However, some concerns may arise in practice. For example, since not all the alternatives are available to the respondents in the real mode choice problem, the observed behavior might not be the same as the hypothetical one.

Another point, which might be more essential, relating to this issue, is that the valuation may be systematically affected by some particular type of choice set. For example, the valuation could be different between public and individual transport and also need not be the same between commuting, and leisure trips.

One of our aims in this study is to investigate whether there is substantial difference in valuation of the types of choice models, that is, between mode choice and within mode choice in our analysis, and also to see if there is a systematic effect in different domains of choice set, such as public versus individual transport, or commuting versus other trips.

3.4 Heterogeneity: observed and unobserved

Numerous studies on the value of time (a summary study, Wardman 2001) have shown that a great deal of variations of estimated values is originated from trip and individual characteristics. In general, there are two ways to take these sources of variations into account in the modeling approach. One is to specify them as the observable variables in the model and the other is by randomizing the parameters or allowing more general correlated error structure form. While the former is referred to '*observed heterogeneity*', the later is regarded as '*unobserved heterogeneity*' in the literature (Brownstone and Small, 2003).

The observed heterogeneity in the estimates can be evaluated by incorporating the interaction terms of those trip or individual traits variables with travel time, reliability, or cost variables. Whilst the idea of testing whether the specification of interaction terms have important effects in valuation seems plausible, our data do not allow us to do this task. There are at least two difficulties. First, for most of the studies we had only very limited information with respect to the estimates. Yet our variables of interest--reliability and scheduling variables ratios—have to be computed from the marginal rate of substitution between travel time and reliability variables. Thus, if there is any traits variable interacting with one of our targeted variables, we are not able to compute these marginal rates of substitutions unless we have further information of some statistics of those traits variables. The second difficulty is that almost no study included the same set of traits variables. Because each study had its own interest and purpose in exploring this issue, therefore, this causes another obstacle to compare those estimated values with each other.

More recent studies have taken the unobserved heterogeneity into account, thanks to the advances in econometrics modeling techniques and computing power. In the literature (Hensher 2001, Greene and Hensher 2003), there are two considerations to accommodate the unobserved variability of preferences into the model: (a) allowing correlation structures of error terms (b) randomizing the parameters associated with each attribute. Nevertheless, it is less clear whether incorporating unobserved heterogeneity will lead to under- or overestimated values. Hensher (2001) suggested that the less restrictive choice model tends to produce higher estimates; while Ghosh (2001) showed that the most general model yielded the lowest estimates, which contradicts Hensher's results.

We aim to investigate the effect of accounting for unobserved heterogeneity on the variables--reliability and scheduling ratios. Because different degrees of complexity were specified in each study to take account of the unobserved variability, it is hard to categorize according to its levels of randomizing parameters or sophisticated error structures. Thus, we only consider the effect on estimates with or without accommodating unobserved heterogeneity.

3.5 Different measurement in attributes

There are various measurements of reliability in empirical assessments, such as standard deviation, coefficient of variation, difference between 90^{th} and medium of travel time etc. Table 2 gives the summary of these different measurements in reliability estimates. This lack of consensus on how to characterize the reliability by a common variable creates some problem in comparison of empirical estimates and this issue will be discussed in more detail in Section 5.2.

In addition to the wide range of reliability measurements, travel time is also evaluated at different grounds, such as mean or medium travel time, free flow time, congested time, and medium delay time, etc. (see Table 3). Since the value of time is the denominator of reliability ratio, these different VOT measurements in travel time may have influence on our variable of interest. In particular, previous studies indicated that value of congested time is considerably higher than value of free flow time or uncongested time (Hendrickson and Plank, 1984) and delay time is evaluated higher than in-vehicle travel time (Wardman, 2001). Thus the different attributes of *VOT unit* may be the main source of reliability ratio variation. However, if we want to classify each VOT attribute into different categories, we would have small sample sizes because of the various uses in specific travel time measurements. In order to solve this problem, we combine some conceptually similarly travel time units into the same class, and distinguish 'congested time and mean delay time' versus 'other'.

4 Methodology

4.1 Data and sampling

To search the empirical estimates for reliability and scheduling variables, we started from the EconLit database, transportation research journals and the google search engine, including published papers, reports, and working papers. Since our variables of interest is the reliability ratio, scheduling ratios, we only considered empirical studies that include the valuation of either both of travel time and reliability or both of travel time and scheduling variables. We computed the reliability and scheduling ratios as the procedure we explained in the end of Section 2.3. However, we excluded some estimates, which used diverging definitions of reliability and cannot be made comparable to other estimates (e.g., Koning and Axhausen 2002[3], Rietveld et al. 2001[4]). The overall studies and computed ratios are shown in Table 1.

4.2 Correction of reliability estimates

As we mentioned in the section 3.5, there are various measurements of reliability and these different uses of reliability measurement certainly create some create in comparison (see Table 2). If we estimate the utility function as Eq (6) for a given set of observations, the coefficient of standard deviation, denoted as STD, and coefficient of coefficient of variation, denoted as CV, cannot be equal, i.e., $\beta_1 \neq \beta_2$ in Eq.(8). The ideal way to correct these coefficients based on different measurements is to go back to the original survey data, and then estimate the model again by using a standard definition of reliability. However, this is not feasible. The second best way to adjust these coefficients is by looking at the relationship between those different measurements then correct the coefficients according to these transformed relationships.

$$U = \alpha \cdot E(T) + \beta_1 \cdot (STD) + ... = \alpha \cdot E(T) + \beta_2 \cdot (CV) + ... \qquad (8)$$

Take STD and CV for example (see Eq (8)), we know in advance that there exists a relationship between STD and CV, that is, $CV = SDT / (mean\ travel\ time)$. Thus, we can infer that $\beta_2 = \beta_1 \times (mean\ travel\ time)$.

Next, we can investigate the relations between standard deviation (STD), difference between 90th and medium travel time (90DMP), and difference between 80th and medium travel time (80DMP) under three types of distributions. Before we look at the relationships between these measurements, we have to make some underlying assumption about the type of distribution of travel time. In the case of uniform distribution, we can derive the analytical solutions for the relations between STD, 90DMP and 80DMP. This shows that that the values of 90DMP and 80DMPare just the scale of standard deviation. Thus, assuming that travel time follows uniform distribution, we can correct the estimated coefficient of 90DMP to standard deviation, based on the calculated ratio. A similar situation holds also for triangle distribution. For the normal distribution, since the analytical solution is difficult to implement, we used simulations to infer these ratios.

[3] Koning and Axhausen used two separate variables 'duration of delay' and 'probability of delay' to present the effect of reliability.
[4] Rietveld et al. defined 'Unreliability' as 15 minutes delay with 50% probability.

The "transformation ratios" are listed in Table 4 for these three distributions. From Table 4 we found out that the values of transformation ratios of normal distribution are laid in between the values of uniform and triangle distributions. Therefore, we decided to choose the transformation ratios for the normal distribution as our "corrected reference". We therefore hypothesize that the distribution of travel time is normally distributed, and then correct the reliability estimates to make them to be comparable.

This correction approach described above can be used in correcting estimates between SDE, CV, 90DMP and 80DMP. Unfortunately, we cannot proceed the same exercise to 'uncertainty' and 'incident' cases. Thus, we will drop those reliability estimates associated with 'uncertainty' and 'incident' variables from our meta-analysis in the next section.

4.3 Overview of empirical estimates

A starting point of meta-analysis is to compare the means of estimates, which are computed from various treatments of categories (e.g., RR in SP and RP studies). The conditional means of RR, SDER, and SDLR on those potential variation factors discussed in the previous section are given in Table 5. Serving as the preliminary stage of meta-analysis, these conditional means give a rough idea about how these factors affect the variables that we are interested in. As we can see from Table 5, these conditional means vary significantly in most of the within-group comparisons, except for the 'unobserved heterogeneity' and 'VOT unit' in RR estimates and 'Lateness variable' in SDER estimates. Note that our RR estimates only vary in 'data types', 'unobserved heterogeneity' and 'utility specification' because of lacking estimates from other dimensions[5]. Though it is only possible to investigate the variation of RR in these three factors, as commented by Brownstone and Small (2003), these three factors are probably the most important sources of variation of VOR estimates except for some observed heterogeneity.

Regarding the direction of the influences in the group comparisons, some striking results can be found in Table 5. First, almost all conditional means vary in a systematical way with respect to those explanatory factors. For instance, the conditional means of the ratios are significantly larger for RP than for SP in all cases. The same statement applies

[5] The VOR estimates for meta-analysis are all evaluated in within mode choice, road transport, and commuting trip. One study, Hensher (2000), is excluded because of the difficulty in comparison caused by adopting different units in reliability estimation.

to other comparison of explanatory factors. Secondly, some of the directions of biases are confirmed with our expectation, for example, including both reliability and scheduling variables induce smaller estimates on both variables. Though it remains unclear for other sources of variation factors, we can explore the data more profoundly in the meta-regression of the next section.

5.0 Empirical results of meta-regression

To explain the variation in reliability and scheduling ratios in a systematical way, we employ the meta-regression technique to meet our purpose. In brief, meta-regression is based on the following relation (Stanley and Jarrell, 1989):

$$y = f(p, x, r, t, l) + \varepsilon,$$

where y is an effect size observed in a series of studies, p is the specific causes, x is moderator variables affecting the cause-effect relationship, and r, t, and l are moderator variables representing differences in research designs, time-periods considered, and locations covered by the initial studies.

In the context of the current analysis, we have three distinct series of effect sizes—reliability ratios, scheduling delay early ratios, and scheduling delay late ratios, as the dependent variables in our OLS regression model. We specify the explanatory variables as the possible causes of variation, and with this specification we basically aim to investigate the effects that we have discussed in the previous section in a multivariate setting. We also consider the time trend and the dummy of US studies to explain the temporal difference and location effect, respectively. The results of regression are reported in the following.

Tables 6-8 show the results of the meta-regression of RR, SDER, and SDLR, respectively. The included sets of explanatory variables are aimed to investigate those sources of variation discussed in Section 3. Yet not all the factors can be investigated in our present study owing to some drawbacks in our database. The small sample size problem and lack of variations in some explanatory variables impose some restriction on including the explanatory variables that we were inclined to, such as trip purpose and location variables. Nevertheless, the large part of the main sources of variations in RR, SDER and SDLR estimates still can be investigated in this framework.

The meta-regression results in Table 6-8 are explained in the following subsections.

5.1 Data types

The results in Table 7 indicate that SP has no significant effect on RR estimates in our meta-regression; whereas Brownstone and Small (2003) concluded that SP underestimated VOT and VOR substantially. A possible explanation for this point is that the SP may underestimate both VOT and VOR in a systematic but equal-proportional way. As a result, this downward bias effect is cancelled out by taking the ratio of these two.

Different from the case of RR, the results obtained from SDER and SDLR show that SP has a highly significant negative effect. The result is quite robust since the conditional means of SDER and SDLR also show the same pattern that SP has lower estimates. One possible explanation for this phenomenon may be the existence of misperception of the amount of schedule delay and the risk aversion behaviour of travellers. The idea is following.

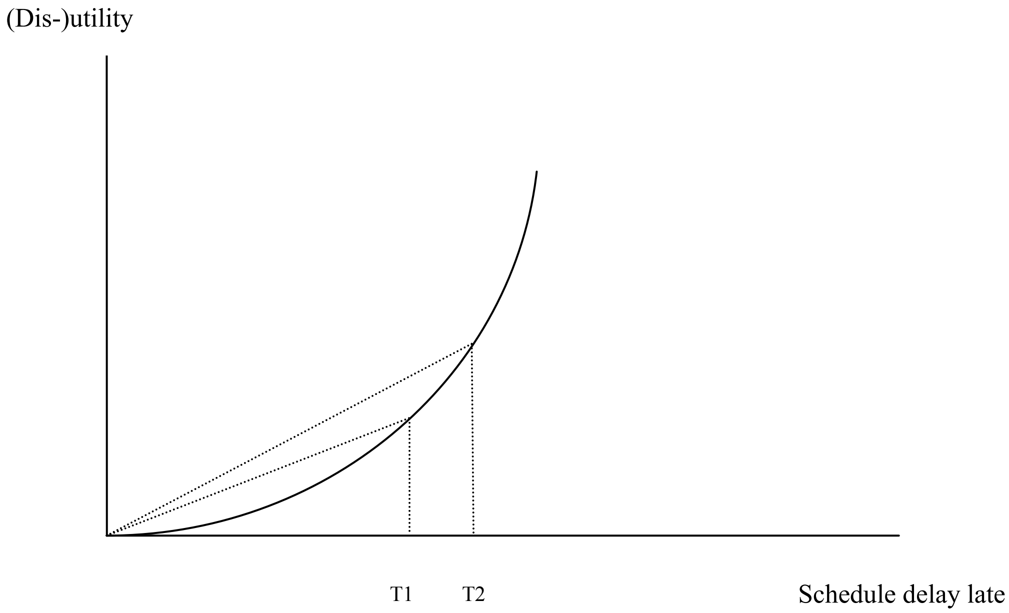

Figure 2 The shape of utility function with respect to schedule delay late variable

If a traveller is risk adverse to SDL, then we can expect that the shape of utility function is convex with respect to SDL (see Figure 2). When a traveller experiences an actual amount of schedule delay late T1, but perceives it as T2, then he may evaluate the value of schedule delay late at T2 instead of the true value T1. From the figure we know that the slope is steeper at T2 than at T1. Thus, the value of schedule delay late may be overestimated under the RP data. Similarly, the value of SDE can be explained in the same manner.

In such a case, if the level of risk aversion is more in schedule delay variables than in travel time, then we can infer that this overestimation of VSDE and VSDL is stronger than of VOT. As a consequence, the difference in risk aversion to schedule delay and travel time may help us to understand why RP could overestimate the ratio of schedule delay.

5.2 Utility specification

Here the utility specification means whether the reliability/scheduling and lateness variables are included in estimated model in studies. In the analysis of reliability ratio, the explanatory dummies 'SCHEDULE' and 'LATENESS' denote the inclusion of scheduling variables and lateness variables in the same estimation model, respectively. Whereas the analyses of schedule delay ratios, we use the explanatory dummy 'RELIABILTY' to indicate the existence of reliability variables in the same estimation model.

The result of including both reliability and schedule variables suggests that there is a significant negative effect on RR estimates as well as on SDER and SDLR estimates. As what we discussed before, the concept of reliability and scheduling delay is not easy to be distinguished and statistically they are positively correlated with each other. Consequently, this negative effect on estimates between each other can be expected if the design of questionnaire was not well specified with respect to these two terms.

The coefficients of 'LATENESS' dummy explain that the inclusion of the lateness variable has a strongly significant negative effect on SDLR estimates; while this effect is not clear in neither RR nor SDER estimates. Since the lateness variable is positively correlated to the SDL variable, the valuation of SDL may be underestimated severely by specifying the lateness variable in the model. This phenomenon is similar to the case of including both reliability and scheduling variables. Our meta-analysis finding is robust in this case and consistent with the general expectation.

5.3 Choice set

In the conditional means analysis, the between mode choice set have higher mean values than within mode choice set. Since we suspect that the between mode choice type may produce a more problematic dataset than within mode choice type because of the sample selection bias, it is interesting to see what the result given by meta-regression is. Though the result shows there is positive significant effect on SDER estimates, it does not give SDLR any significant outcome. Given our rather small and less robust database, it is improper to draw the conclusion that between mode choice set would certainly produce higher estimates of SDER and SDEL than within mode choice set. Whether between mode choice set has generally higher estimated values than within mode choice set remains further investigation.

5.4 Unobserved heterogeneity

Unlike the result in conditional means, the meta-regression result shows that accounting for unobserved heterogeneity has a negative impact on our three dependent variables, though this effect is only significant on SDER estimates. Since the direction of effect in meta-regression is opposite to what we found in conditional means, it is less clear which direction of effect should be the true one. Actually, with different degrees of complexity and different types of specification of accommodating the unobserved heterogeneity into the model, the result is probably mixed. Whether the consideration of unobserved heterogeneity has certain effects on empirical estimates requires further information on modeling details and richer database.

5.5 Different measurement attributes

In the investigation of different measurements of travel time, we find out that there is a negative effect if the value of time is evaluated at the congested time. The effect is highly significant for SDER and SDLR. This result corresponds to our anticipation since the congested value of time is higher in general, and hence the computed ratios should have small values.

6.0 Conclusions

Since the last decade, reliability and scheduling delay of travel time are considered as important factors in traveler's decision making. Many researchers have attempted to model the reliability and scheduling delay attributes into traveler's choice model. As a result, a wide range of estimated values is produced owing to the different data types or methodologies used in the valuation. Our aim in this present paper is to analyze the explanatory factors that systematically affect our variables of interest—reliability ratio

(RR), scheduling delay early ratio (SDER), and scheduling delay late ratio (SDLR) by means of the multivariate statistical technique: meta-analysis.

We start by correcting the reliability estimates that evaluated under different measurements, i.e. coefficient of variation, standard deviation, difference between 90th and medium, and difference between 80th and medium of travel time. After making these reliability estimates to be comparable, we use several multivariate regression models to further explore the sources of variations among empirical estimations in RR, SDER, and SDLR. Explanatory variables included in our meta-analysis are the type preference data, the choice type, the trip mode, different VOT unit measurements, the inclusion of schedule and reliability attributes, and the inclusion of lateness attributes.

We find that, as expected, the inclusion of both reliability and scheduling attributes (SDE, SDL) would lead to lower estimated values for both attributes. A similar result is also found in the case of SDL and lateness variable. Regarding the types of data, a striking finding is that the SP data may produce lower values for SDER and SDLR than the RP data. The misperception error of the magnitude together with the risk aversion attitude associated with schedule delay late/early variables may be one of the possible explanations. Still, to obtain more robust evidence for the understating problem of SP we need more empirical studies to confirm.

Our analysis raises the interesting question whether the valuation of reliability or scheduling variables should be based on within mode choice or between mode choice type questions. Though the result shows that between mode choice type has higher estimates in SDER, it does not provide the same evidence in RR and SDLR estimates. We can only suspect that between mode choice type question may create more variation in empirical valuations due to the sample selection bias.

It remains unclear that whether accounting for unobserved heterogeneity has significant influence on RR, SDER, and SDLR estimates in our meta-analysis. Nevertheless, we believe that accounting unobserved behavior heterogeneity, e.g. nested correlations among choice alternatives, more general error structure forms, or unobserved random effects in individuals (randomizing the parameters associated with some attributes) etc., in a more sophisticated manner will result to more accurate estimates and this is what future researches should head to.

Table 1 Overview of studies with empirical estimates of reliability ratio, schedule delay early ratio and schedule delay late ratio

Authors	Study type	Year of Publication	VOR ratio (RR)		VSDE ratio (SDER)		VSDL ratio (SDLR)	
			obs	mean	obs	mean	obs	mean
Small	RP	1982	-	-	2	0,667	2	2,139
Wilson(89)	RP	1989	-	-	4	4,742	4	5,888
Lam and Small	RP	2001	17	1,062	2	0,326	2	0,562
Small et al	SP	1999	2	2,303	-	-	-	-
Ghosh (Dissertation)	SP& RP	2001	5	0,986	-	-	-	-
J. Yan (Dissertation)	SP& RP	2002	30	1,082	-	-	-	-
Noland (Dissertation)	SP	1995	3	0,536	4	0,872	4	1,813
Koskenoja (Dissertation)	SP	1996	7	0,378	7	0,507	5	1,396
Bates et al	SP	2001	-	-	1	0,442	1	0,897
Hensher	SP	2001	6	0,750	-	-	-	-
A. de Palma	SP	2003	-	-	5	0,454	5	1,780
G. de Jong et al.	SP	2003	-	-	8	1,020	8	1,409
Cascetta and Papola	RP	2003	-	-	5	2,301	5	4,392

Table 2 Different reliability unit attributes used in empirical estimations

Unit attributes	Notation	# obs	Min	Max	Mean
Standard deviation of travel time	STD	4	0.548	3.222	1.140
Coefficient of variation of travel time	CV	8	0.131	0.576	0.357
Difference between 90th and 50th travel time	90DMP	20	0.483	1.714	0.925
Difference between 80th and 50th travel time	80DMP	19	0.968	1.952	1.469
Incident	INC	11	0.380	0.441	0.421
Uncertainty	UNC	6	0.541	1.461	0.750

Table 3 Different travel time unit attributes used in empirical estimations (with adjusted VOR ratios)

Unit attributes	Notation	VOR ratio_adjust				VSDE ratio				VSDL ratio			
		#obs	Min	Max	Mean	#obs	Min	Max	Mean	#obs	Min	Max	Mean
Travel time	TT	30	0.10	2.51	0.85	25	0.23	2.92	1.13	24	0.57	5.88	2.20
Free flow time	FF	-	-	-	-	2	0.31	0.59	0.45	2	1.96	2.42	2.19
Congested time	CT	16	0.48	1.71	0.88	5	0.21	0.68	0.40	4	0.37	1.44	0.97
Medium time savings	MTS	5	0.43	1.32	0.96	-	-	-	-	-	-	-	-
Mean delay	MD	-	-	-	-	1	0.44	0.44	0.44	1	0.90	0.90	0.90

Table 4 Transformation ratios between STD, 90DMP, and 80DMP for various distributions

	Uniform	Normal	Triangle
STD	1,000	1,000	1,000
90DMP	1,384	1,283	0,993
80DMP	1,038	0,843	0,661

Table 5 Conditional means of VOR ratios (RR), VSDE ratios (SDER), and VSDL ratios (SDLR) for various categories of studies

	VOR ratio studies (n=51)		VSDE ratio studies (n=33)		VSDL ratio studies (n=31)	
Groups	n	mean	n	Mean	n	Mean
Data Types						
Revealed preference	38	0.9477**	9	1.4991***	9	3.0405***
Stated preference	13	0.6375**	23	0.7583***	21	1.5230***
Choice types						
Between mode choice	-	-	13	1.5128***	13	2.5563**
Within choice mode	-	-	19	0.5930***	17	1.5445**
Trip mode						
Private transport	-	-	17	0.7636**	16	1.5322**
Public transport	-	-	15	1.1967**	14	2.4982**
Trip purpose						
Commute	-	-	17	0.7509**	16	1.5375**
Others	-	-	15	1.2111**	14	2.4921**
Unobserved Heterogeneity						
Not account for	41	0.8548	15	0.7947***	14	1.7093
Unobserved hetero.	10	0.9253	17	1.1183***	16	2.2224
VOT unit attributes						
Congested travel time	30	0.8458	26	1.0950**	25	2.1889**
Otherwise	21	0.9011	6	0.4103**	5	0.9537**
Utility specification I						
No scheduling/reliability variable	41	0.9702***	23	1.1113**	22	2.2436**
Including scheduling / reliability variable	10	0.4522***	9	0.5968**	8	1.2662**
Utility specification II						
No lateness variable	43	0.9116*	15	1.1252	15	2.6013***
Including lateness variable	8	0.6373*	17	0.8266	15	1.3647***

Note: The statistical test (t-test) is concerned with the comparison of means within each group. Significance is indicated by ***, **, and *, referring to significance at the 1%, 5%, and 10% levels.

Table 6 Results of meta-regression of reliability ratio (RR)

Categories	Variables	OLS	OLS with robust SD	WLS	WLS with robust SD
Fixed effect	Constant	0.625	0.625	20.654*	20.654
		(0.31)	(0.21)	(1.89)	(1.31)
Data type	SP	0.002	0.002	0.222	0.222
		(0.01)	(0.01)	(0.74)	(0.53)
Choice type	BETWEEN	-	-	-	-
		-	-	-	-
Unobserved heterogeneity	HET	-0.075	-0.075	0.137	0.137
		(-0.48)	(-0.67)	(0.67)	(0.59)
Trip mode	CAR	-	-	-	-
		-	-	-	-
VOT unit	VOT_CT	-0.056	-0.056	-0.309*	-0.309
		(-0.39)	(-0.42)	(-1.72)	(-1.55)
Time trend	YEAR	0.014	0.014	-0.086	-0.086
		(0.20)	(0.13)	(-1.24)	(-1.07)
Utility Specification I	SCHEDULE / RELIABILITY	-0.572**	-0.572	-0.724***	-0.724***
		(-2.24)	(-1.00)	(-4.53)	(-4.82)
Utility Specification II	LATENESS	0.117	0.117	0.004	0.004
		(0.57)	(0.96)	(0.04)	(0.06)
R-squared		0.2391	0.2334	0.4291	0.4291
Adj R-squared		0.1353	-	0.3362	-
Probability value F-test		0.0508	-	0.0006	-
Number of observations		51	51	51	51

Note:
1. Significance is indicated by ***, **, and *, referring to significance at the 1%, 5%, and 10% levels, respectively, with t-values in parentheses.
2. "Robust SD" means the OLS or WLS is estimated with robust standard errors, in such a case we specify the structure of error terms in OLS and WLS, which is correlated within studies but uncorrelated between studies.
3. "WLS" means weighted least squared where the weight corresponding to each study is computed by the square root of the sample sizes of each study.[6]

[6] Ideally, the weights should be based on the accuracy of the parameters in estimation. However, it is not feasible in our case owing to lack of information. Theoretically, the variance is inversely related to the sample size, we use the sample sizes in original studies to calculate the appropriate weights.

Table 7 Results of meta-regression of schedule delay early ratio (SDER)

Categories	Variables	OLS	OLS with robust SD	WLS	WLS with robust SD
Fixed effect	Constant	0.988***	0.988**	-9.238	-9.238
		(2.79)	(3.40)	(-0.85)	(-0.53)
Data type	SP	-0.694**	-0.694*	-1.247***	-1.247**
		(-2.68)	(-2.10)	(-3.46)	(-2.95)
Choice type	BETWEEN	0.768**	0.768*	0.594*	0.594
		(2.32)	(2.10)	(1.92)	(1.60)
Unobserved heterogeneity	HET	-0.540**	-0.540***	-0.721**	-0.721***
		(-2.08)	(-3.53)	(-2.50)	(-3.79)
Trip mode	CAR	0.112	0.112	0.152	0.152
		(0.59)	(0.69)	(0.73)	(1.22)
VOT unit	VOT_CT	-0.560***	-0.560	-0.310	-0.310
		(-2.02)	(-1.36)	(-1.30)	(-1.18)
Time trend	YEAR	0.031**	0.031**	0.089***	0.089***
		(1.41)	(1.25)	(4.90)	(4.51)
Utility Specification I	SCHEDULE / RELIABILITY	-0.340	-0.340**	-0.283	-0.283**
		(-1.29)	(-1.29)	(-1.52)	(-3.27)
Utility Specification II	LATENESS	-0.387**	-0.387	0.071	0.071
		(-2.10)	(-1.44)	(0.45)	(0.81)
R-squared		0.7684	0.7684	0.6441	0.6441
Adj R-squared		0.6879	-	0.5203	-
Probability value F-test		0.0000	-	0.0009	-
Number of observations		32	32	32	32

Note:
1. Significance is indicated by ***, **, and *, referring to significance at the 1%, 5%, and 10% levels, respectively, with t-values in parentheses.
2. "Robust SD" means the OLS or WLS is estimated with robust standard errors, in such a case we specify the structure of error terms in OLS and WLS, which is correlated within studies but uncorrelated between studies.
3. "WLS" means weighted least squared where the weight corresponding to each study is computed by the square root of the sample sizes of each study.

Table 8 Results of meta-regression of schedule delay late ratio (SDLR)

Categories	Variables	OLS	OLS with robust SD	WLS	WLS with robust SD
Fixed effect	Constant	3.215***	3.215***	21.044	21.044
		(5.28)	(9.66)	(0.96)	(0.63)
Data type	SP	-1.722***	-1.722***	-2.285***	-2.285**
		(-3.87)	(-5.08)	(-3.21)	(-2.75)
Choice type	BETWEEN	0.285	0.285	-0.482	-0.482
		(0.50)	(0.81)	(-0.80)	(-0.68)
Unobserved heterogeneity	HET	-0.191	-0.191	-0.376	-0.376
		(-0.42)	(-0.97)	(-0.67)	(-1.15)
Trip mode	CAR	0.379	0.379**	0.581	0.581***
		(1.14)	(2.64)	(1.39)	(4.14)
VOT unit	VOT_CT	-1.500***	-1.500***	-1.681***	-1.681**
		(-2.95)	(-3.89)	(-3.30)	(-2.84)
Time trend	YEAR	0.034	0.034	0.154***	0.154***
		(0.90)	(1.27)	(4.30)	(3.78)
Utility Specification I	SCHEDULE / RELIABILITY	-0.848*	-0.848**	-0.600	-0.600**
		(-1.89)	(-2.64)	(-1.62)	(-2.94)
Utility Specification II	LATENESS	-1.297***	-1.297***	-0.793**	-0.793*
		(-4.13)	(-4.07)	(-2.36)	(-2.19)
R-squared		0.8097	0.8097	0.6975	0.6975
Adj R-squared		0.7371	-	0.5822	
Probability value F-test		0.0000	-	0.0004	
Number of observations		30	30	30	30

Note:
1. Significance is indicated by ***, **, and *, referring to significance at the 1%, 5%, and 10% levels, respectively, with t-values in parentheses.
2. "Robust SD" means the OLS or WLS is estimated with robust standard errors, in such a case we specify the structure of error terms in OLS and WLS, which is correlated within studies but uncorrelated between studies.
3. "WLS" means weighted least squared where the weight corresponding to each study is computed by the square root of the sample sizes of each study.

References

Arup Transport Planning, Modelling and appraisal of journey time variability, Report for UK Department for Transport (DfT). May, 2002.

Bates J., Polak J., Johes P., Cook A., 2001, The valuation of reliability for personal travel, Transportation Research Part E37, pp. 191-229.

Bates J., Fearon J., and Black I., Frameworks for modeling the variability of journey times on the highway network: a report for UK DfT, Dec., 2003.

Ben-Akiva M., Lerman S.R., 1997, Discrete choice analysis, The MIT Press, Cambridge.

Brownstone D., Small K. A., 2003, Valuing time and reliability: assessing the evidence from road pricing demonstrations, University of California, Irvine, Working paper.

Cascetta E., Papola A., 2003, A joint mode-transit service choice model incorporating the effect of regional transport service timetables, Transportation Research Part B37, pp. 595-614.

Ghosh A., 2001, Valuing time and reliability: commuters' mode choice from a real time congestion pricing experiment, PhD Dissertation, University of California, Irvine.

Hendrickson C., Plank E., 1984, The flexibility of departure times for work trips, Transport Research Part A 18 (1), pp.25-36.

Hensher David A., 2001, The valuation of commuter travel time savings for car drivers: evaluating alternative model specification, Transportation 28, pp. 101-118.

Jackson W. B., Jucker J. V., 1981, An empirical study of travel time variability and travel choice behavior, Transportation Science Vol. 16, Iss. 4, pp. 460-475.

Jong G. de, Daly A., Pieters M., Vellay C., Bradley M., Hofman F., 2003, A model for time of day and mode choice using error components logit, Transportation Research Part E39, pp. 245-268.

Koskenoja P. M. K., 1996, The effect of unreliable commuting time on commuter preferences, PhD Dissertation, Department of Economics, University of California, Irvine.

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

Louviere J. J., Hensher D. A., Swait J. D., 2000, Stated choice methods, Cambridge University Press, Cambridge.

Noland R. B., Polak J. W., 2002, Travel time variability: a review of theoretical and empirical issues, Transport Review Vol. 22, Iss. 1, pp. 39-54.

Noland R. B., Small K. A., 1995, Travel time uncertainty, departure time and the cost of the morning commute, Paper presented to the 74th annual meeting of the Transportation Research Board, Washington, Jan. 1995.

Palma A. de, 2002, Departure time choice models estimates for Paris area, Presentation at DTLR Workshop 2002.

Rietveld P., Bruinsma F.R., Vuuren D.J. van, 2001, Coping with unreliability in public transport chains: a case study for Netherlands, Transportation Research Part A 35, pp. 539-559.

Small K. A., 1982, The scheduling of consumer activities: work trips, American Economic Review 72, pp. 467-479.

Small K. A., Noland R., Chu X., Lewis D., 1999, Valuaiton of travel-time savings and predictability in congested conditions for highway user-cost estimation, Transportation Research Board NCHRP Report 431.

Small K. A., Noland R. B. Koskenoja P., 1995, Socio-economic attributes and impacts of travel reliability: a stated preference approach, University of California, Irvince, California PATH research report.

Small K. A., Winston C., Yan J., 2002, Uncovering the distribution of motorsits' preferences for travel time and reliability: implications for road pricing, University of California Transportation Center, working paper #546.

Stanley T.D., Jarrell S.B., 1989, Meta-regression analysis: a quantitative method of literature surveys**,** Journal of Economic Surveys 3, pp. 161-170.

Wardman M., 2001, A review of British evidence on time and service quality valuations, Transportation Research Part E 37, pp. 107-128.

Wilson P. W., 1989, Scheduling costs and the value of travel time, Urban Studies 26, pp. 356-366.

Yan J., 2002, Heterogeneity in motorists' preferences for travel time and time reliability: empirical finding from multiple survey data sets and its policy implications, PhD Dissertation, Department of Economics, University of California, Irvine.