TRAIN PASSENGERS' VALUATION OF TRAVEL TIME UNRELIABILITY

Maria Börjesson
Centre for Transport Studies - WSP Analysis & Strategy
maria.borjesson@wspgroup.se

Jonas Eliasson
Centre for Transport Studies - Royal Institute of Technology
jonas.eliasson@infra.kth.se

2008-11-22

Presented at the 2008 European Transport Conference, Leiden, Netherlands.

Abstract
Investments in rail infrastructure are often motivated by the need to reduce travel time and to reduce travel time unreliability. Knowing travellers' valuation of travel time unreliability relative to in-vehicle time and travel cost is hence important for cost-benefit analysis of rail investments. For scheduled services with relatively high reliability and long headways, the most common approach seems to be the "expected delay" approach, where passengers are assumed to value a possible delay proportional to its expected value.

Using three different data sets, we investigate whether the expected delay approach holds empirically. In particular, we study how the valuation of a possible delay with probability *p* and length *L* depends on *p* and *L*.

The main result is that this valuation is *not* proportional to the expected delay, but increases slower than linear in the delay probability. This is particularly pronounced for small risks. This means that estimated "values of delay time" will depend on the delay risk level *p*: "delay time values" will be higher the lower the risk levels in the study are. This means that estimated values of delay times which do not take the non-linearity in delay risk into account will result in valuations that cannot be transferred between contexts with different delay risks. We also give a theoretical reason why this should be an expected phenomenon, as long as headways are large.

Regarding the dependence on delay length *L*, the valuation increases slower than linearly in L in about half of the cases, whereas in the remaining cases, the linearity hypothesis (the valuation is proportional to delay length) cannot be rejected. Contrary to what might be expected, the valuation never increases faster than linearly in delay length.

Accounting for random population heterogeneity turns out be important. In particular, the valuation of delays is shown to have a much larger variation across the population than the valuation of travel time or travel cost.

1 INTRODUCTION

The purpose of this paper is to investigate how passengers on long-distance trains value unexpected delays relative to scheduled travel time and travel cost. We use three different data sets to explore how the value of a possible delay varies with the delay risk and delay length. In particular, we are interested in investigating whether the commonly used "expected delay" approach holds empirically.

Investments in rail infrastructure are often motivated by the need to reduce travel time and to reduce travel time unreliability. Knowing travellers' valuation of travel time unreliability relative to in-vehicle time and travel cost is hence important for cost-benefit analysis of rail investments. Several countries (to our knowledge Holland, Sweden and the UK) have introduced travel time reliability benefits in their official appraisal schemes (CBA guidelines etc.). Knowledge of travellers' valuation of delays is also crucial for the construction of (socially) efficient timetables, and for the possibility of constructing useful incentive contracts between train operators and track operators. Such incentive contracts have recently been tested by the Swedish Rail Administration and the main Swedish train operator SJ.

The variability of travel time in general has received increasing interest the last few years, both in terms of research and applications. Most studies of travellers' valuation of travel time variability have used one of two possible different approaches: the scheduling approach, where the traveller's departure time choice is explicit in the model, or the reduced-form approach, where some measure of the variability is introduced directly in a reduced-form indirect utility function. As a measure of the variability, most studies have used either the standard deviation of the travel time or the expected delay compared to scheduled arrival time, although some studies include both (Batley et al., 2007) and some studies use percentiles of the travel time distribution (Lam and Small, 2001).

In the literature, the standard deviation approach seems to be the method of choice for situations with high travel time variability and good opportunities for travellers to adjust their departure time. Typical settings would be congested car traffic and urban (high-frequency) transit services. The expected delay approach, on the other hand, seems to be the method of choice for scheduled services with fairly high reliability (i.e. where most arrivals are "on time") and limited flexibility of departure time due to long headways. A typical example is long-distance train trips. For example, this is the measure used for Swedish train CBA (SIKA, 2008), and it is the recommended measure in the UK Passenger Demand Forecasting Handbook.

Although this generalisation is obviously not clear-cut, there are two good reasons for the measures of variability to be different in the two types of context. First, it can be proved that when departure time can be chosen freely (and a few other assumptions), the resulting disutility of travel time variability will be proportional to the standard deviation (see references and further explanations in section 2). If departure time is fixed, on the other hand, one can derive (under certain fairly restrictive assumptions) that the resulting disutility of travel time variability will be proportional to the expected delay (see section 2). Second, the travel time variability for car and urban transit trips is typically much higher than for long-distance trains[1], and this has implications for the way stated preference surveys present different variability levels to respondents. While in high-variability contexts, the method of choice seems to be to illustrate the travel time distribution by presenting a representative sample of travel times (typically 5-10), the method of choice for longer train trips seems to be to state the risk of delay of certain length(s), compared to the scheduled arrival time. Examples of the former presentation method can be found in Black and Towriss (1994), Small et al (1999), Bates et al (2001) and Hollander (2005a), while examples of the latter method can be found in Kroes and Kouwenhoven (2005, 2007), Rietveld et al. (2001) and Eliasson (2004).

As stated above, the focus of this study is to investigate how passengers value a risk for a delay with a given probability and length. In particular, we are interested in the whether the disutility of such a delay risk is proportional to its expected value. This is not only important for cost-benefit analysis: it is also commonly used as a quality measure by train operators. For example, both UK and Swedish train operators use “average delay per train” as an indicator of punctuality. Since it can be expected that the choice of quality indicator also affects which measures and investments are chosen, knowing whether this is a good quality measure has practical importance.

Section 2 presents a brief overview of relevant theory and literature. In the final part of the section3, we expand and discuss the motivation for the use of the expected delay approach. Section 3 presents the data material, and section 4 presents estimation results. Section 5 concludes.

2 THEORY, TERMINOLOGY, LITERATURE

As stated above, travel time variability can be (roughly) grouped into types: scheduling models and reduced-form utility models. In the former type, the traveller’s choice of an optimal margin to reduce consequences of a possible delay is modelled explicitly. In the latter type, some measure of travel time variability is introduced in the (indirect) utility function. The two most common variability measures are the standard deviation of travel time and the expected delay compared to scheduled arrival time. Below, we will first discuss the scheduling model and how it leads (under certain assumptions) to either the

[1] The coefficient of variation of car travel times in congested conditions is typically 0.2-0.3 (Black and Towriss, 1993; INCA, xx; Eliasson, 2005), which also seems to be a representative figure for transit trips (although the coefficient of variation for transit trips is harder to define and compare, since it will depend on the service frequency and the number of interchanges): it is for example consistent with available Stockholm data on delays and average trip times. This can be contrasted with long-distance train trips, where typically 80-90% of trains are ”on time” (usually defined as within 3 or 5 minutes within scheduled arrival time; Rietveld, 2001; Kroes and Kouwenhoven, 2007; see also below), implying a coefficient of variation in the range 0.05-0.15.

standard deviation approach (treated first) or the expected delay approach (later in the section).

2.1 Scheduling models

Following Vickrey (1969) and Small (1982), assume that the traveller's utility depends on the travel time and the deviation from the planned arrival time. Note that the *planned* arrival time is not necessarily the same as the *preferred* arrival time: this often creates confusion. In short, the preferred arrival time is the time where traveller would like to arrive, while the planned arrival time is the time when the traveller plans to arrive. There are several reasons why these might be different: for example, the traveller may choose a different departure time to avoid congestion, or because transit headways put constraints on the departure time. (A further discussion can be found in Börjesson, 2006, together with a model which both contain the scheduling cost of deviation from the preferred arrival time and the cost of travel time variability, i.e. the risk for deviation from the planned arrival time.)

Using the notation from Fosgerau and Karlström (2007), the scheduling utility function is typically some variation of

$$u(D,T) = \alpha D + \omega T + \beta(T-D)^{+} \qquad (1.)$$

where T is the random travel time and D is the headstart relative to the planned arrival time. The notation $(x)^{+}$ is a shorthand for the function $\max(x,0)$. ω is the direct utility of time spent on the trip, and α is the (relative) opportunity cost of travel time ("time as a resource" to use a term from De Serpa, 1971). α is sometimes referred to as "the disutility of being early", but we prefer to view it as the utility difference between time spent at the origin and time spent at the destination before the planned arrival time – hence, the (relative) opportunity cost of time. β is the disutility of "lateness", i.e. arrival after the planned arrival time. Usually, the "lateness penalty" is assumed to be linear in the lateness, although some authors have suggested a step function (an extra penalty as soon as the traveller arrives after the planned arrival time), and some authors have suggested that the marginal disutility of lateness may be either increasing or decreasing.

Several studies have estimated variants of such utility functions, resulting in what is called scheduling models. Given the utility function and the distribution of the travel time, travellers' optimal headstart D* can (in principle) be calculated, and hence the disutility incurred by the travel time uncertainty. This will result in a reduced-form utility function $u^*(\Phi_T) = E_T[u(D^*,T)]$, taking only the travel time distribution Φ_T as argument. An advantage of scheduling models is that they can also serve as departure time models. Their main weaknesses are that they need information of travellers' preferred arrival times, and are often computationally intensive to use.

2.2 Mean-variance models

Often, the effect on travel time variability on departure times is of less interest. Instead, the purpose of a study may be to obtain a value of travel time variability for use in cost-benefit analysis or demand modelling. In such cases, estimating the reduced-form utility function u* directly is often the preferred method, in particular since this circumvents the need to solve for D* and then calculating u* in a second step. Various forms of such reduced-form utility functions have been proposed and used, the most common being the mean-variance approach (where the mean and the standard deviation of Φ_T enter the utility function) and the expected delay approach (where the scheduled travel time and the average delay enter the utility function).

Fosgerau and Karlström (2007), extending results by Noland and Small (1995) and Bates et al. (2001), show that as long as the headstart D can be chosen freely and that the travel time distribution is independent of departure time (although they show that this condition is of less practical importance), the reduced form utility function can be written as

$$u^* = (\alpha+\omega)t + \beta H(\alpha/\beta,\Phi')\sigma \qquad (2.)$$

where Φ' is the standardised distribution of the travel time T, t the average travel time E(T) and σ the standard deviation of T. H is a functional depending only on the tail of the standardised travel time distribution and the parameter ratio α/β. An important conclusion is that the reduced-form utility function can always be written as a linear combination of the mean travel time and the standard deviation of travel time. Another important conclusion is that the parameter in front of the standard deviation will depend on the standardised travel time distribution. Standard deviation valuations that have been obtained in one context can thus not be directly transferred to another context with a different standardised travel time distribution.

Hence, as long as the traveller can choose departure time freely, the mean-variance approach is well motivated for theoretical reasons. It also seems to be the most common approach in studies of non-scheduled trips – car trips and urban (high-frequency) transit trips. Examples of mean-variance studies are abundant, e.g. Jackson and Jucker (1982), Black and Towriss (1993), Senna (1994) and Eliasson (2004). A good survey of results from both mean-variance and expected delay studies can be found in Vincent (2008). Mean-variance studies are often compared with each other using the *reliability ratio*, the ratio between the standard deviation parameter and the travel time parameter, i.e. $(\alpha+\omega)/\beta H$ in formula (2). Typical reliability ratios are in the range 0.5 – 1.3 (see Hollander, 2005b).

2.3 Expected delay models

For low-frequency scheduled trips, however, the expected delay approach seems to be most common. Examples of such studies are abundant, e.g. Hensher and Prioni (2002), Algers et al. (1995), Rietveld et al. (2001), Kroes and Kouwenhoven (2005), Eliasson (2004) and Kroes et al. (2007). Wardman (2001) presents a meta-analysis of British valuations of expected delay (among other things). A good survey of both mean-variance and expected delay studies can be found in Vincent (2008). It also appears to be a fairly common practice to use the expected delay as a quality parameter used for empirical quality evaluations; for example, the Swedish Rail Administration uses expected delay together with the share of trains "on time" as their main punctuality indicators, as does the UK train companies (Batley et al., 2007). However, as we will show below, one needs to make further rather restrictive assumption for the approach to be valid.

Usually, expected delay studies present the distribution of delays to the respondent in the form “there is a risk p that the train is delayed L minutes”, occasionally using several delay length L_i and corresponding risk levels p_i. Denoting the scheduled travel time by T_o and the random delay by Δ, the utility function is then assumed to be on the form (assuming the more general although unusual form with several risks p_i and lengths L_i)

$$u^* = (\alpha+\omega)T_0 + (\beta+\omega)E(\Delta) = (\alpha+\omega)T_0 + (\beta+\omega)\Sigma_i p_i L_i \qquad (3.)$$

Expected delay studies are often compared with each other using the *reliability multiplier* (a term introduced (?) by Batley et al, 2007), defined as the ratio between the expected delay parameter and the travel time parameter, i.e. $(\alpha+\omega)/\beta$ in formula (3).

The theoretical and empirical foundations of the expected delay approach, however, are fairly weak. Note that the value of a delay of length L and risk p is supposed to be linear in both L and p. Both assumptions are contestable on theoretical grounds, and testable empirically. Empirical investigations are presented in section 4. Before the empirical investigations, let us consider how we would expect the valuation of a delay *L* with risk *p* should depend on *L* and *p*. Assume that we can write this valuation as

$$u = (\alpha+\omega)T_0 + (\beta+\omega)g(p)f(L) \qquad (4.)$$

Then, the question is what the functions g() and f() look like. The expected delay approach assumes that both g() and f() are linear.

As long as the departure time is fixed, the expected value of the delay is obviously linear in the delay risk p (assuming the underlying model above is correct). But if travellers have at least some possibility to adjust their departure time, even if the frequency is low, then this will not hold. The easiest way to show what happens is to assume that departure time may be chosen freely, and to use the simplistic but frequently used assumption that delays occur with frequency *p* and length *L*. The utility function (1) can then be written as

$$u = (\alpha+\omega)T_0 + (\beta+\omega)pL \qquad (5.)$$

It is easy to see that if $(\alpha+\omega) < (\beta+\omega)p$, then the traveller should choose a departure time T_0+L minutes before the preferred arrival time. This will guarantee that she arrives at or before the preferred time, incurring a generalised travel $(\alpha+\omega)(T_0+L)$. If instead $(\alpha+\omega) > (\beta+\omega)p$, then the traveller should choose a departure time T_0 minutes before the preferred arrival time. This will mean that she will suffer delays of length L with frequency p, incurring a expected generalised travel of $(\alpha+\omega)T_0 + (\beta+\omega)pL$. This means that travellers will choose either of two extremes: they either depart early enough that they are always in time, or they depart so they will be on time if there is no delay.

But naturally, all travellers are not equal. Some will be have a high valuation of late arrival (large β), and some a low valuation. It is easy to show that this means that the "reliability multiplier", the ratio between the travel time parameter and the expected-delay parameter, will be decreasing in *p*. In other words, the reliability multiplier should be higher for small delay risks. Intuitively, this is because the higher the delay risk is, the more passengers will have sufficient margins so a delay will not make them arrive after their preferred arrival time (and hence suffer a lateness penalty). For very high delay risks, the reliability multiplier will simply tend to the value of travel time $\alpha+\omega$. Hence, we would expect that g() increases slower than linearly in p,

Turning to f(L), a natural conjecture would be that f(L) would increase faster than linearly in L. The reason would be that since the timetable constrains the choice of departure time, most passengers will have a margin between the train's scheduled arrival time and their actual preferred arrival time, after which a delay penalty is incurred. This would mean that the longer the delay, the more passengers will arrive after their preferred arrival time. In our notation, this would mean that for small L, when most passengers have a margin larger than L, the delay valuation will be close to ω (the direct utility of travel time), while for large L, when most passengers have a margin less than L, the delay valuation will be close to $\omega+\beta$.

But there are also possible reasons why the valuation might increase slower than linearly in L. For example, one can hypothesize that for some travellers, once an appointment or a connection has been missed, being even later does not increase the disutility much – the damage is already done. Hence, the behaviour of f() is ambiguous: if the delay penalty is (at least) linear, we would expect f() to increase faster than linearly in L (since more and more people will be delayed more than their margin), whereas f() might increase slower than linearly if the delay penalty increases slower than linearly.

3 THE DATA

Our data material consists of three data sets from two surveys – one survey contained two different stated choice games, with different question types. All studies described the risk for late arrival as (some variation of) "x out y trains are z minutes late; the rest are on time". That is, rather than illustrating a smooth distribution with e.g. a sample of travel times, delays were presented as "with probability *p*, there is a delay of length *L*".

The first survey was conducted in May 2004 on trains between Stockholm and Gothenburg. Respondents answered 8 pairwise choices, where the variables were travel time, travel cost, delay risk and delay length. Below is an example of a pairwise choice.

Choose one!	Departure 1	Departure 2

Fare	**20** SEK more than today	As today
Delays	**1** out of 10 trains are **20** minutes late, the rest are on time	**3** out of 10 trains are **10** minutes late, the rest are on time
Travel time	The scheduled travel time is **2 hours 30 minutes**	The scheduled travel time is **2 hours 45 minutes**
I prefer:	☐ **1**	☐ **2**

The second survey was conducted in May 2007 on the train between Stockholm and Norrköping (many of these trains then continue to Gothenburg). This survey contained two stated choice exercises. In the first, respondents answered 8 pairwise choices, where the variables were travel cost, delay length at a given delay risk (always the same risk in both alternatives), whether passengers received information about the length of a delay if there was one, and whether passengers received compensation (coffee or a ticket voucher) if there was a delay. Below is an example of a pairwise choice.

Choose one!	**Departure 1**	**Departure 2**
Fare	As today	**70** SEK more than today
Delay length	Once every two weeks, there is a **40 minutes** delay	Once every two weeks, there is a **10 minutes** delay
Compensation	None	Coffee
Information:	Should there be a delay, you get information about delay length and possible connections.	Should there be a delay, you get information about delay length and possible connections.
I prefer:	☐ **1**	☐ **2**

Note that the delay risk was always the same in both alternatives, although the risk was different in different pairwise choices. The pilot studies had indicated that questions that included both delay risks and lengths (and the other variables) were too complicated for respondents, so it was decided that delay risk should not enter the pairwise choice explicitly (although the delay risk will of course affect the trade-off between delay, cost and compensation/information). In the 2004 study, using different risk levels in a pairwise choice seemed to have worked, though, but then there were fewer other variables. A closer study of the 2004 study also reveals that the random distribution of the parameters is much larger than in the 2007 study, which could depend on the questions being more difficult.

In the second exercise, a novel question type was used, called *Most Preferred Improvement* (MPI) (see Levander, 2007). The respondent is faced with a number of potential improvements of her trip, and is asked which one she would prefer. The presented improvements related to the fare, travel time, delay length, delay risk, information and compensation. All improvements are relative to a (partly hypothetical) "current situation" with the actual fare, delays on average once a week and 30 minutes long, and no compensation or information should a delay occur. Below is an example of an MPI choice.

Which of these improvements would you prefer?	
☐	**30** SEK lower fare
☐	The train is delayed on average **once every two weeks** instead of once every week
☐	If a delay occurs, it is on average **15** minutes instead of 30 minutes
☐	You get **information** about the delay length and connections, should a delay occur.

The table below summarises facts about the three data sets.

Abbreviation	**PC04**	**PC07**	**MPI07**
Date	May 2004	May 2007	May 2007
Type	Pairwise choice	Pairwise choice	Most preferred improvement
No. of respondents (*valid choices*)	402 (*2920*)	2270 (*15471*)	2270 (*7720*)
- commuters	52 (*367*)	539 (*2839*)	539 (*1438*)
- other private	172 (*1221*)	801 (*5316*)	801 (*2615*)
- business	178 (*1332*)	930 (*7316*)	930 (*3667*)
Average fare	562 SEK	449 SEK	449 SEK
Variable ranges			
- delay risk *p*	5%, 15%	10%, 2.5%	20%, 10%, 5%
- delay length *L*	5-55 min.	5-40 min. (at 10% risk), 20-115 min. (at 2.5% risk)	15-30 min.
- fare	-55 to +130 SEK compared to current fare	+10 to +150 SEK compared to current fare	-15 to -40 SEK compared to current fare
- travel time	2:30 – 3:30	n/a	n/a
- information about delay length and connections	n/a	Yes/no	Yes/no
- compensation	n/a	Ticket voucher, coffee, none	Ticket voucher, coffee, none

4 ESTIMATION RESULTS

In the estimations, we will explore how a delay of length L and probability p is valued. We will concentrate on three questions:

- How does the value of the delay depend on delay length L?
- How does the value of the delay depend on delay probability p?
- How is the value of the delay affected by allowing for randomly distributed parameters?

We will use the principal functional form

$$u = \alpha t + (\gamma_1+\gamma_2 Y)c + \beta_p f(L) \tag{6.}$$

t is the scheduled travel time, c the fare, Y is income and β_p a parameter that is different for different values of p and f(L) some function of delay length L. One purpose of the study is to test the "expected lateness approach", i.e. the assumption that f(L) is linear in L and that the β_p:s will vary as $\beta_p = \bar{\beta}*p$. The expression $\beta_p f(L)/\lambda$ can be thought of as the *value of a possible delay* whereas the expression $\beta_p f(L)/(\lambda p)$ is usually called the *value of [expected] delay time*.

The standard deviation approach is seldom (if ever) used when delays are described as "delays have probability p and length L", but it can nevertheless be illuminating to note that the standard deviation approach would imply that f(L) is a linear function and that the β_p:s should vary as $\beta_p = \bar{\beta}\sqrt{p(1-p)} \approx \bar{\beta}\sqrt{p}$ for small[2] p.

For each model structure, nine separate models are estimated, corresponding to the three data sets divided into three trip purposes (commuting, other private and business).

4.1 Is the value proportional to the delay length?

To explore the dependence on the delay value of delay length, we use two different methods. First, we estimate a polynomial of order two:

$$u = \alpha t + (\gamma_1+\gamma_2 Y)c + \beta_p L + \theta_p L^2 \tag{7.}$$

If θ_p is significantly different from zero, then obviously the delay value is not linear in L, and the sign of θ_p will reveal whether it increases slower or faster than linearly. Note that we estimate different θ_p:s for different risk levels p.

Second, we estimate a Box-Cox function:

$$u = \alpha t + (\gamma_1+\gamma_2 Y)c + \beta_p (L^{\lambda_p}-1)/\lambda_p \tag{8.}$$

[2] For example, for p = 0.15 (which would be high delay risk), the error of this approximation is 8%. For p = 0.05, a more typical risk, the error is 3%.

Similarly, the size of λ_p (less or greater than 1) will reveal whether f(L) is linear or not. Note that we estimate different λ_p:s for different risk levels p.

Full estimation results are given in appendix. The table below summarises the results, focusing on whether f(L) is approximately linear or not. The shorthands "$\theta_p \approx 0$" and "$\lambda_p \approx 1$" mean that the parameters were not significantly different from zero and one, respectively, at the 95% level. "n/a" means that the model was not possible to estimate.

It should be noted that the models discussed in the table below do not include any treatment of the 'repeated measures' property of SP data (there are no individual level error components). Now, experience suggests that this usually lead to an overstatement of the significance of the parameters, which should be borne in mind when interpreting the results. "n/a" means that the model did not give meaningful results, for example because of insignificant delay parameters.

Data set	**Trip purpose**	**Box-Cox: $\lambda_p \neq 1$?**	**Quadr. pol.: $\theta_p \neq 0$?**
PC04	All	n/a for p = 5% λ_p = 0.4 for p = 15%	$\theta_p \approx 0$ for p = 5% $\theta_p > 0$ for p = 15%
	Commuting	n/a	$\theta_p \approx 0$ for p = 5% $\theta_p > 0$ for p = 15%
	Other private	n/a	$\theta_p \approx 0$ for p = 5% $\theta_p > 0$ for p = 15%
	Business	n/a	$\theta_p \approx 0$ for p = 5% $\theta_p > 0$ for p = 15%
PC07	Commuting	$\lambda_p \approx 1$	$\theta_p \approx 0$
	Other private	$\lambda_p \approx 1$	$\theta_p \approx 0$
	Business	λ_p = 0.5 for p = 2.5% $\lambda_p \approx 1$ for p = 10%	$\theta_p > 0$ for p = 2.5% $\theta_p \approx 0$ for p = 10%
MPI07	Commuting	n/a	$\theta_p > 0$ for p = 20%
	Other private	n/a	$\theta_p > 0$ for p = 20%
	Business	n/a	$\theta_p \approx 0$ for p = 20%

First we note that the conclusions of the Box-Cox models and the polynomial models coincide. To summarise the results, the hypothesis that the delay value is proportional to delay length often holds: In about half of the cases, the linearity hypothesis cannot be rejected. But when the linearity hypothesis *is* rejected (in about half of the cases), it is always the case that the delay value increases *slower* than linearly in L. Contrary to what might have been expected, the delay value never increases faster than linearly in L. The slower-than-linear behaviour mostly occurs at high risk levels (($p \geq 15\%$), while for low risk levels ($p \leq 10\%$), the linearity hypothesis can seldom be rejected. The difference from the linearity assumption at the high risk levels is considerable: the λ parameter is around 0.5, meaning that the delay value increases only proportional to the square root of the delay length.

This result is rather non-intuitive, since we would expect that the longer the delays are, fewer travellers would have a sufficient safety margin, which imply that disutility increase with delay length. However, one hypothesis that might explain that the delay value increases slower than delay length is that some respondents have a low acceptance for high risk levels irrespective of delay time, but higher acceptance for lower risk levels. If this is the case the non-linearity found would be captured by a dummy penalty for the higher risk level:

$$u = \alpha t + (\gamma_1+\gamma_2 Y)c + \beta_p L + \theta\delta \quad (9.)$$

where δ is 1 if the risk level is 0.15 and 0 otherwise. Only the data set PC04 can be used to investigate this explanation, since delay risk did not vary within the pairwise choices in the PC07 survey.

Estimating this model formulation we make two observations. First, the log likelihood value increases compared to the quadratic-polynomial model for all three model segments. Second, only the parameters β_p changes significantly compared to the linear model and the quadratic-polynomial model. So it appears as though the last explanation is valid and that there are at least some travellers that have a low acceptance for high risk levels even for short delays.

4.2 Is the delay value proportional to the risk level?

Next, we turn to the question of how the value of a possible delay with probability p and length L depends on the risk level p. To do this, we estimate nine separate models (three data sets divided into three trip purposes) using the following functional form:

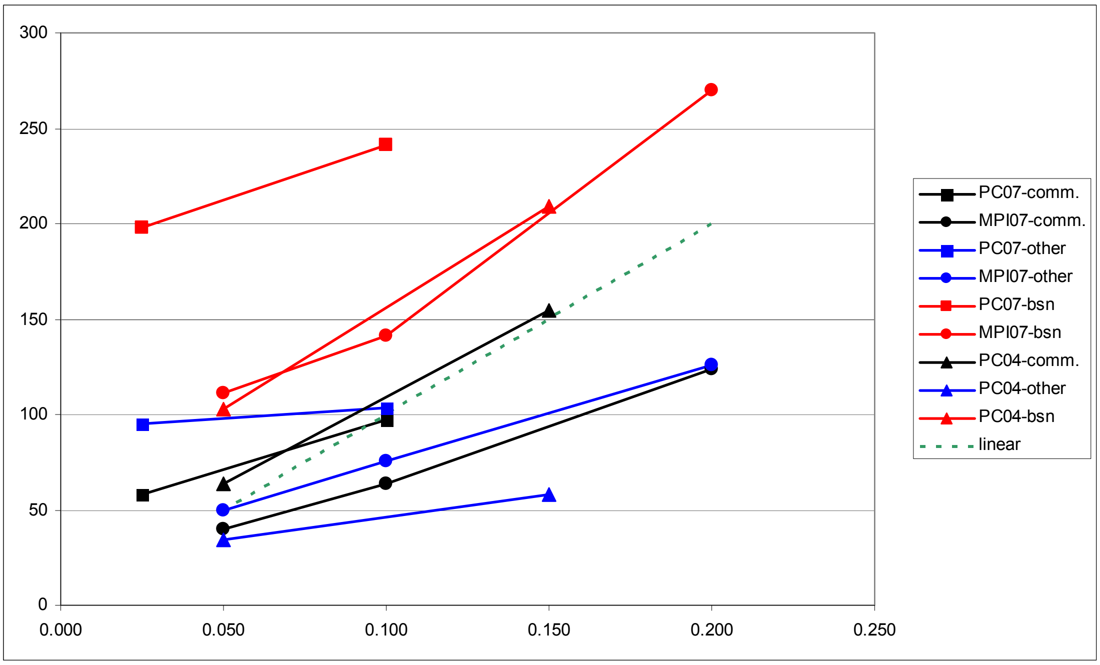
$$u = \alpha t + (\gamma_1+\gamma_2 Y)c + \beta_p L \quad (10.)$$

Hence, we estimate separate β_p:s for different risk levels. Note that the conventional "value of [expected] delay time" would correspond to β_p/p.

These models were estimated as random-parameters logit models, assuming that β_p and γ_1 are normally distributed (other distributions were tested with disappointing results). The 'repeated measures' property was accounted for by assuming that the random parameters (which can also be interpreted as error components) are individual specific. Valuations were evaluated at the estimated means. For the PC07 and MPI07 data sets, this did not change the results, but for the PC04 data set, this turned out to be important: model fit improved very much, and the valuations (which had been high compared to the other studies when using standard multinomial logit) became more in line with the other data sets, and also with what we would have expected from other studies. The main result below – the elasticities of delay values with respect to risk levels – do not change if standard multinomial logit is used, though.

The diagram and table below show the value of a possible delay of one hour for different risk levels (in SEK). Complete estimation results can be found in the appendix.

Risk	Commuting	Other private	Business	Source
2.5%	58	95	198	PC07
5%	40	50	111	MPI07
5%	64	35	103	PC04
10%	97	103	241	PC07
10%	63	75	142	MPI07
15%	155	58	209	PC04
20%	124	126	270	MPI07



To reveal how the delay value depends on the risk level, it is illuminating to compute the arc elasticity of the delay value with respect to the risk level for each study, i.e.

$$\mathrm{VoD} \propto p^{\varepsilon} \Rightarrow \varepsilon = \ln(\mathrm{VoD}_1/\mathrm{VoD}_2)/\ln(p_1/p_2) \qquad (11.)$$

where VoD_i is the value of a possible delay with risk p_i. The PC07 and PC04 contained two risk levels each, and hence give one elasticity each. The MPI07 data set contains three risk levels, and hence give two elasticities. The table below shows the results.

Average risk level $(p_1+p_2)/2$	Commuting	Other private	Business	Source
6.25%	0.4	0.1	0.1	PC07
7.5%	0.7	0.6	0.3	MPI07
10%	0.8	0.5	0.6	PC04
15%	1.0	0.7	0.9	MPI07

A zero elasticity means that the value of a possible delay is completely independent of the risk level. This might be plausible if the traveller feels compelled to *always* have a sufficient margin to compensate for a delay, no matter how small the risk is. The "expected lateness" assumption implies an elasticity of 1: the value of a possible delay is proportional to the risk level. Finally, assuming that the value of a possible delay is proportional to the standard deviation implies an elasticity of 0.5 (for small risks).

From the table, we make the following observations:

1. All elasticities are less or equal to 1, and almost all are strictly less than 1. Hence, the value of a possible delay increases slower than linearly in p – in particular for small risks. Conversely, halving a small risk will not nearly decrease the delay value by half, but with much less.
2. The smaller the risk, the smaller the elasticity; for very small risks, the elasticities for non-commuting trips are almost zero, meaning that the value of a possible delay is almost independent of the risk level.
3. The elasticities for commuting trips are consistently higher than for the other trip purposes. One could hypothesize that the need to always have sufficient margins, even at very small risks, is higher for non-commuting trips. This interpretation is supported in focus groups reported in Kroes et al (2007), who note that "unexpected train delays were often accepted as a valid excuse for arriving late at their work, so part of the disbenefit of being late could be transferred to the employer".

The main message here is that the expected lateness approach is clearly not valid for small risk levels (10% or below). For higher risk levels (15% and above), it might be reasonable approximation. Recall that typical punctualities for train services lie in the range 85-95%. Hence, using the expected lateness approach seems like a bad idea.

As a corollary, this means that the concepts of "the value of delay time" and "the reliability multiplier" have to be used with great caution, since these will depend on the risk level. In principle, these values can be computed and used at a fixed, given risk level – but they cannot be transferred to a context with another risk level.

5 CONCLUSIONS

The purpose of this paper is to investigate how passengers on long-distance trains value unexpected delays relative to scheduled travel time and travel cost. In particular, we study how the valuation of a possible delay with probability p and length L depends on p and L.

The main result is that this valuation is *not* proportional to the expected delay, but increases slower than linear in the delay probability. This is particularly pronounced for small risks.

This means that estimated "values of delay time" will depend on the delay risk level *p*; "delay time values" will be higher the lower the risk levels in the study are. This means that estimated values of delay times which do not take the non-linearity in delay risk into account will result in valuations that cannot be transferred between contexts with different delay risks.

Train operators commonly use "average delay per train" as a quality measure. The findings here indicate that this may not be an optimal measure, if the purpose is to improve the overall attractiveness of the train service.

Regarding the dependence on delay length *L*, the valuation increases slower than linearly in L in about half of the cases, whereas in the remaining cases, the linearity hypothesis (the valuation is proportional to delay length) cannot be rejected. Contrary to what be expected, the valuation never increases faster than linearly in delay length.

Accounting for random population heterogeneity turns out be important. In particular, the valuation of delays is shown to have a much larger variation across the population than the valuation of travel time or travel cost.

6 REFERENCES

Algers, S., Dillen, J., Widlert, S. (1995) The National Swedish Value of Time Study. Swedish Institute for Transport and Communications Analysis (SIKA).

Bates, J., J. Polak, P. Jones, A. Cook (2001) The valuation of reliability for personal travel. Transportation Research 37E, 191-229.

Batley, R., Ibáñez, J.N., Wardman, M., Shires, J. and Whelan, G. (2007) A discrete choice study to assess the impact of reliability on passenger rail demand. Proceedings of the 2007 European Transport Conference.

Black, I.G. and J.G. Towriss (1993) Demand effects of travel time reliability. HMSO.

Börjesson, M. (2008) Joint RP–SP data in a mixed logit analysis of trip timing decisions. *Transportation Research E* 44:6.

De Serpa, A.C. (1971): A theory of the economics of time, The Economic Journal, 81, 828-846.

Eliasson, J. (2004) Car drivers' valuations of travel time variability, unexpected delays and queue driving. Proceedings of the 2004 European Transport Conference.

Fosgerau, M. and Karlström, A. (2007) The value of reliability. Working paper available at http://mpra.ub.uni-muenchen.de/5733.

Hollander, Y (2005a) The attitudes of bus users to travel time variability. Proceedings of the 2005 European Transport Conference.

Hollander, Y (2005b) Direct versus indirect models for the effects of unreliability. Transportation Research A 40 (9), 699-711.

Jackson, W. B. and Jucker, J. V. (1982) An Empirical Study of Travel Time Variability and travel Choice Behavior, Transportation Science, 16:4, 460-475.

Kroes, E. and Kouwenhoven, M. (2005) On the value of punctuality on suburban trains to and from Paris. Proceedings of the 2005 European Transport Conference.

Kroes, E., Kouwenhoven, M., Duchateau, H., Debrincat, L. and Goldberg, J. (2007) The value of punctuality on suburban trains to and from Paris. 2007 TRB annual meeting.

Lam, T. and K. Small (2001) The value of time and reliability: measurement from a value pricing experiment. *Transportation Research* **37E**, 231-251.

Levander, A. (2007) Theory and Experimental Design in Non Market Valuation. Application to Value of Time with Empirical Assessment and Monte Carlo Evidence Master's Degree Project Stockholm, Sweden 2006.TEC-MT-07-003.

Noland, R.B. and Small, K. (1995) Travel time uncertainty, departure time and the cost of the morning commute. Transportation Research Record 1493, 150–158.
Noland, R. and J. Polak (2002) Travel time variability: a review of theoretical and empirical issues. Transport Reviews 22-1, 39-54.
Rietveld, P., F.R. Bruinsma and D.J. van Vuuren (2001) Coping with unreliability in public transport chains: A case study for Netherlands. Transportation Research 37E, 539-559.
Senna, L. A. D. S. (1994), The Influence of Travel Time Variability on the Value of Time, Transportation, 21, pp. 203-228.
Small, K. A., Noland, R., Chu, X. & Lewis, D. (1999), Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation, NCHRP Report 431, Transportation Research Board, U.S.
Wardman, M. (2001) A review of British evidence on time and service quality valuations. Transportation Research 37E, 107-128.
Vickrey, W.S., (1969) Congestion theory and transport investment. *American Economic Review*, Papers and Proceedings 59:251-260.
Vincent, M. (2008) Measurement Valuation of Public Transport Reliability. Booz Allen Hamilton. Available at www.landtransport.govt.nz/research/reports/339.pdf

