Multimodal Travel Time Variability
Final Report

11\textsuperscript{th} November 2008

JOHN BATES SERVICES

Richard Batley, Susan Grant-Muller, John Nellthorp, Gerard de Jong, David Watling (ITS Leeds)
John Bates
Stephane Hess and John Polak (CTS Imperial)

A report for the Department of Transport
Multimodal travel time variability

VERSION Final
18th November 2008

Richard Batley
Susan Grant-Muller
John Nellthorp
Gerard de Jong
David Watling
Institute for Transport Studies,
University of Leeds,
Leeds, LS2 9JT
Tel: +44 (0)113 343 5325
Fax: +44 (0)113 343 5334

John Bates
John Bates Services
1 Picklers Hill, Abingdon OX14 2BA
Tel: +44 (0)1235 526702

Stephane Hess
John Polak
Centre for Transport Studies
Imperial College London
London SW7 2AZ
Tel: +44 (0)20-7594-6100
Fax: +44 (0)20-7594-6102
Contents

Executive Summary .................................................................................................................. 4
1. Introduction .......................................................................................................................... 8
   1.1 Outline of the study ........................................................................................................ 8
   1.2 Definitions of terms ...................................................................................................... 11
   1.3 Sources and Measurement of Travel Time Variability .............................................. 12
   1.4 Requirements for modelling and appraisal ................................................................. 13
2. Theoretical methods for demand modelling and valuation of variability ................. 15
   2.1 Choice under Uncertainty ......................................................................................... 15
   2.2 The mean vs. variance approach ............................................................................... 20
   2.3 The scheduling approach ......................................................................................... 22
   2.4 Equivalence between the mean vs. variance approach and scheduling approach . 23
   2.5 Applying the theory to the valuation of reliability ................................................... 24
   2.6 Challenges and opportunities for further theoretical research ................................ 25
   2.7 Recommendations on theoretical approach .............................................................. 28
3. Evidence on Valuation of Travel Time Variability ....................................................... 30
   3.1 Introduction .................................................................................................................. 30
   3.2 Presentational issues relating to variability (passengers) ......................................... 31
   3.3 Derivation of PAT for scheduling models (passengers) ........................................... 37
   3.4 Issues in valuation studies for freight transport ....................................................... 38
   3.5 Empirical findings for passenger and freight ............................................................ 39
   3.6 Recommendations on valuation ............................................................................... 41
4. Supply issues ....................................................................................................................... 47
   4.1 Highway assignment models: state-of-practice and issues for representing TTV .. 48
   4.2 Representing actual TTV in highway assignment models ....................................... 51
   4.3 Modelling highway network performance under TTV ............................................. 54
   4.4 Towards ‘Implementable’ Models ............................................................................. 56
   4.5 Widening the Range of Models ............................................................................... 63
   4.6 Public transport supply issues .................................................................................. 65
   4.7 Concluding remarks ................................................................................................. 66
5. Integration with Appraisal ................................................................................................. 68
   5.1 Introduction .................................................................................................................. 68
   5.2 Implications for incorporating “Reliability” into the Appraisal Procedure ............ 69
6 Conclusions .......................................................................................................................... 70
Annex I: Development of theory ......................................................................................... 72
Annex II: Valuation of travel time reliability ......................................................................... 79
Annex III: Arup Transport planning Appendix B .................................................................. 88
Annex IV: References ............................................................................................................ 100
Executive Summary

This report has been commissioned by the Department for Transport under contract number PPRO 4/3/5 as a scoping study with the aims of a) defining a way forward on the theory of (travel time) reliability across modes; and b) identifying areas of research needed to arrive at a practical solution to quantifying and valuing (travel time) reliability changes. A consideration of terminology concluded that the terms reliability and variability should not be used interchangeably because their meanings are opposites despite common practice. Hence the focus has been defined as ‘travel time variability’, one possible measure being the variance or standard deviation of travel time.

The underlying theory is outlined according to the basic principles and evolution of the theory in recent research. The two major candidates are the so-called “mean-variance” approach, and the “schedule utility” approach, each employing rather different utility functions. For the first it is assumed the traveller will consider travel time variability by choosing the departure time that maximises his or her expected utility, where expected utility could be approximated as a linearly additive function of the mean and variance of the travel time distribution, though in practice the standard deviation of travel time (having the same units as the mean) is often substituted for the variance: confusingly, both cases are referred to as “mean-variance” approaches. Empirical investigations have frequently simplified from the full functional form and with the assumption of constant coefficient of variation this has led to the so-called ‘reliability ratio’ term. This is given by the ratio of the coefficient of the standard deviation to the coefficient of the mean within the expected utility function. Variability can then be approximately allowed for simply by a “loading” for the value of time using this ratio.

The second basic approach is based upon the notion of scheduling and reformulates the utility function for a particular departure time choice as a function of four components; travel time, ‘schedule delay early’ (SDE), ‘schedule delay late’ (SDL), and a ‘lateness’ dummy variable (unity if schedule delay late is non-zero). The latter three components are conditioned by the notion of a ‘Preferred Arrival Time’ (PAT). Journeys arriving before the PAT are deemed ‘early’: the SDE is the difference between the PAT and the actual arrival time, SDL is zero. Journeys arriving after the PAT are deemed ‘late’: the SDL is the difference between the actual arrival time and the PAT, the lateness dummy variable is unity and SDL is zero. These four components of the utility function are typically specified as linearly additive. Variability in travel time is accommodated by taking expectations of each component of utility over the travel time distribution giving rise to a term representing the probability of lateness.

Work has been undertaken to investigate the equivalence between the two approaches at both a theoretical and empirical level. The evidence from both perspectives suggests this is highly dependent on the travel time distribution. Any equivalence between the two approaches also depends on the ability of travellers to continuously vary their departure time, which is generally not the case for public transport. In our view, it is not currently possible to choose between these two approaches without a clearer understanding of their
similarities and differences. While the mean vs. variance approach, together with the associated ‘reliability ratio’, has the advantage of greater simplicity, this does not detract from our belief that the scheduling approach may reveal important, and unique insights into behavioural responses to unreliability, particularly in the context of scheduled (but infrequent) public transport services. Specifically, further research is needed in the following areas: translating the theory to public transport, the interpretation of lateness and alignment with industry conventions, valuing travel time variability (as distinct from valuing time risk) and the extension of choice dimensions.

Empirical evidence on the valuation of travel time variability is then presented drawing on research evidence for passenger and freight journeys from the UK and abroad. Almost all the empirical work to obtain values for variability, using either the mean vs. variance or the scheduling approach or both, has been based on Stated Preference (SP) data. This is due to the difficulties in collecting Revealed Preference (RP) data that includes measures of variability, travel time and travel costs that are not heavily correlated. The study results are determined both by the way data has been collected and by the way it has been analysed, resulting in a wide range of studies with outcomes which are not necessarily presented in a comparable format. Where comparable results are given there is still diversity in the empirical values obtained due to the nature of the study, the sample taken and the level of disaggregation in the findings. While all reviewed studies agree that variability is a factor of substantial importance, there are no generally accepted monetary values for variability, or indeed a reliable estimate of the relative weight of travel time and travel time variability. A recommendation from this study is for substantive new research into the valuation of variability, investigating both how information on variability is presented to respondents and how it is later specified in econometric models estimated from the data. Before this can take place, agreement must to be reached on the theoretical framework for journey time variability as outlined above.

For the purposes of this study, the review of supply effects concerning TTV has been largely confined to the role of the network model. Although the body of research has addressed the performance of both public and private transport networks, the majority of the work has focused on the highway context. Two broad areas are considered here: firstly, how to represent actual TTV, and secondly, how to model the effect of TTV on network performance. On the first issue the approaches may be divided into those (the majority) that look at the composite effect on TTV of all sources, and those that aim to decompose the variability into its component sources, ie day-to-day-variability of demand and incidents. Many studies have focused on the standard deviation as sufficient for representing TTV, but there is a question as to whether the complete distribution of travel times is required to obtain a true picture of TTV impacts. Further thought should be given to whether the objective of representing actual TTV needs to be widened to summary measures beyond the second moment, given the typically asymmetric impact of flow variations and incidents on the right-hand tail of the travel time distribution.

A number of theoretical and practical challenges remain in incorporating travel time reliability in network assignment models, echoing the debate on the relative merits of the mean vs. variance and scheduling approaches. A more fundamental question is also
raised, ie what the mechanism for transferring advances in network assignment theory into practice is. Looking to widening the scope of modelling tools considered for this purpose, the methods illustrated as examples are practicable and realistic. While they do not go so far as to criticise the theoretical foundation of current practice (e.g. they accept concepts such as ‘equilibrium’ even those these would not be accepted by all researchers when modelling a variable environment), they do need specialist software tools that can make the best use of current research and knowledge. The incorporation of travel time reliability within practical transport planning tools is an area that is evolving quickly and requires the continued support of the Department to keep apace with the research developments.

Against this background it is challenging to make recommendations on procedures for use in multi-modal appraisal. There is a practical imperative, however as failing to analyse variability implies omitting variability improvements from appraisal and decision-making, in direct contradiction both to policy (for example, the Eddington Study) and the majority of the research evidence. The evidence indicates that transport users are willing-to-pay something for improvements in travel time variability – the questions are really 'how much' and 'for which measure of variability'. A judgement is therefore needed on whether the theory and evidence is sufficient to propose an interim approach to evaluating variability. As a response to the practical imperative the DfT have recently produced a new consultation version of TAG Unit 3.5.7, recommending using the best available data/knowledge available for practitioners at this time. Essentially this implies use of mean-delay approach for rail and the reliability ratio for other modes.

We understand the need for a possible interim approach. Nonetheless, our general view is that proceeding along these lines is premature until a) we have agreed on a theoretical formulation, and b) resolved what practical models can be built. At this stage we have therefore gone no further than outlining a list of the elements of the appraisal that will need to be renovated once both the theoretical formulation and practical models are in place.

A key recommendation from this research is that there is a need for a major new study into the valuations of variability, investigating both the question of how information on variability is presented to respondents and how it is later specified in econometric models that are estimated on the data. This must, however, follow on from agreement on the theoretical framework for journey time variability. The aim should be to achieve a high level of consistency between the underlying theory, the data collected and the estimated model.

With respect to incorporating travel time reliability in network assignment models, a number of theoretical and practical challenges remain, mainly echoing the issues concerning the relative merits of the mean vs. variance and scheduling approaches. A more fundamental issue remains, however: namely, the mechanism for transferring advances in network assignment theory into practice. For further advances in the practice of TTV modelling to take place, the practice of issuing guidance for use in a small
number of accepted commercial packages may need to be reconsidered - the use of specialist software tools that can make the best use of current research and knowledge should also be supported by the Department.
1. Introduction

1.1 Outline of the study

This report has been commissioned by the Department for Transport under contract number PPRO 4/3/5 as a scoping study to address the issue of general topic of (travel time) reliability with the following aims:

- to define an agreed way forward on the theory of (travel time) reliability across modes;
- to suggest areas of research by which DfT could expect to arrive at a practical solution to quantifying and valuing (travel time) reliability changes.

Although the term “reliability” is widely used, we have been concerned as a team that it is too broad and ill-defined a term to be useful. In the next sub-section, we discuss general questions of terminology, as a pre-requisite to setting out the theoretical issues. However, in the light of that discussion, we have decided to use the term “travel time variability” in preference, and henceforth our discussion will be in those terms. This has also necessitated a change in the title of this report.

The range of transport interventions that the Department would wish to model or would wish scheme promoters to model is broad, and therefore the theoretical approach should be general enough to be potentially transferable to all of them. Whilst passenger trips are the primary focus, the role of freight in theoretical and empirical evidence is summarised, with key findings and references included. Moreover the objectives of the Scoping Study are twofold. Firstly, to review and summarise the current state of the art; covering the theory, evidence, appraisal, valuation and tools for travel time variability in a multimodal context. Secondly, to outline recommendations on research to take the state of the art forward towards a set of technically defensible but practical and operational recommendations on the measurement and inclusion of travel time variability impacts in new scheme appraisal.

A summary of the structure to the work is given in Figure 1.1 – this is a complex area of research and part of the challenge of the research is to provide a consistent and integrated approach to the treatment of variability based on a number of previous studies which have taken place on a more ad-hoc basis.

The first stage of this study has been to review and summarise the current state of knowledge on the modelling and valuation of travel time variability for surface transport. The underlying theory is outlined in section 2, describing the basic principles of dealing with travel time variability and evolution of the theory in recent research. Following on from this, empirical evidence on the valuation of travel time variability is presented in section 3, drawing on the research evidence for passenger and freight journeys from the UK and abroad. As the number of studies is relatively small, we summarise each one in terms of the key features of the method and the subsequent outcomes, and we identify gaps in knowledge in travel time variability valuation. In section 4 a discussion of how
travel time variability could be included in practical network modelling is given. Two broad areas are considered here: firstly, how to represent actual travel time variability and secondly how to model the effect of travel time variability on network performance. The focus does not extend to microsimulation models and remains largely on highway network performance due to the well developed nature of the research in that field. Finally, in section 5, the implications of the above for appraisal are outlined. Obviously, the ability to carry out appraisal is dependent upon being able to measure and forecast journey time variability, which in itself is dependent on the state of the art in theory and valuation.

Whilst the underlying objective in each section has been to aim towards a consistent way forward, areas of difference in academic perspective are also represented. During the process of gathering the theoretical and practical evidence a number of gaps in theory, evidence and methods have been identified. These form the basis for areas of further research which have been outlined within each section.
Chapter 2: State of the art on theory:
theoretical models based on:
  a) Mean-variance of journey time
  b) Scheduling

Chapter 3: Evidence base for value of variability:

Passenger transport: scheduling approach evidence
Passenger transport: mean-variance approach evidence
Freight evidence base

Issues in integration with Network models

Chapter 4: Supply Issues
Incorporating reliability in supply models

Needs for practical software tools

Chapter 5: Integration with Appraisal
Implications for Integration with the appraisal and presentation of the results

Figure 1.1: Structure of the research
1.2 Definitions of terms

It is essential to be clear about terminology from the start. We will use the commonly accepted scientific definitions for variance and standard deviation, as follows:

- **variance**, the mean of the squared deviations of a set of quantities, in this context journey times;
- **standard deviation**, the square root of the variance.

We will use a definition of **variability** from the Oxford English Dictionary:

- **variability**, the fact of, or capacity for, varying in amount, magnitude or value;
- hence **journey time variability**, the fact of, or capacity for, variation in journey time.

Therefore:

- **variance of travel time or standard deviation of travel time** are measures of the variability of journey time;
- other measures which could be used to characterise the variability of journey time include measures of skewness or kurtosis, although these measures are only meaningful if variance is non-zero;
- knowledge of the whole probability distribution of journey times would give the most comprehensive understanding of variability.

**Reliability** itself has been the most challenging to define; however, the OED helps us to see that there are multiple definitions:

- **reliability**, the quality of being reliable – **reliable**, that may be relied upon; in which reliance or confidence may be put; trustworthy, safe, sure;
- **reliability (Statistics)**, the extent to which a measurement made repeatedly in identical circumstances will yield concordant results.

The second, narrower definition of reliability seems to imply the reverse of **variability**, i.e. a journey time which does not vary when repeated. We have said that variability can be measured by the standard deviation or variance, and it seems that perfect reliability corresponds to zero variance. In this report we aim to simplify by using the term **variability** for this narrower concept.

The first, broader definition includes the ideas of trustworthiness and reliance, in this context whether a transport service can be trusted and relied upon by users, in terms of its journey time. The determinants of reliability in this broader sense are likely to include the **information** available to users, and the ways in which users form their **expectations** about journey time. For example, if a timetabled transport service is on average 10 minutes late

---

compared with the timetable, whether or not this is a problem for users will depend their
capacity to anticipate it. Users who rely on the timetable information – irregular users, for
example – may find their expectations are wrong and they are late at their destination.
Regular users of the service may form different expectations, learning by experience, and
in effect modifying or rejecting the timetable information.

These issues are not limited to public transport, since irregular users will also face
challenges predicting car journey times on the road network, even with the help of
journey planning tools. These are under-researched areas of the topic of reliability, in
which it turns out to be harder to make recommendations for modelling and appraisal
practice, and where we are working on ideas for relevant research.

In the light of these definitions, it would be preferable not to use the terms reliability and
variability interchangeably, both because that risks confusing a broader with a narrower
interpretation of reliability, and because their meanings are opposites, the former
suggesting low variance, the latter suggesting high variance.

On this basis, as already noted, we prefer not to use the term reliability, which we will
leave to be used for wider concepts. We re-define our subject matter as relating to travel
time variability, and we note that one possible measure is the variance or standard
deviation of travel time. Nevertheless, the most appropriate measure is not being defined
at this stage, for reasons which we now go on to discuss.

1.3 Sources and Measurement of Travel Time Variability

Having clarified the general definition, there remain problems of application and
measurement, and these are largely related to the source and context of travel time
variability. For example, the journey time between two stations on the rail network may
show appreciable variation due to a different pattern of services (eg fast trains on the
hour, slow trains on the half-hour). These are scheduled differences, and are in principle
completely predictable (ie timetabled). Hence, even though some travellers may not be
aware of them, they can be ascertained, and should therefore not be included in a
measure of variability.

In a similar vein, on the highway side there are observable regularities in the variation of
travel time by time of day (eg peak vs off-peak), and day of week. Much of this “regular”
variation will be due to the impact of demand variation, through its influence on
congestion. These variations are not scheduled, as in the rail service pattern case
discussed above, but they are, at least to some extent, predictable. As Bates (2000) has
put it:

Since the essence of any measure of variability (such as the variance) relates the
variations to the expected value, alternative definitions of the expected value will clearly
have an impact. A failure to clarify this point in the past has led to much confusion of

12
measurement. In general, it is sensible to remove as far as possible any non-random effects.

For example, a person who regularly travelled by car into the city centre in the early hours of the morning might be surprised by the time that the same journey took when he made it in the morning rush hour, but it would be perverse to view this as an example of unreliability. ...... it is reasonable to expect travellers to foresee the impact of predictable variations in demand. Such impact relates predominantly to the highway mode, via the usual supply (speed-flow) relationships, but there are also potential impacts on rail modes. From a modeller’s point of view, there are, in addition to variations in the (average) demand profile during the course of a day, seasonal effects, day-of-week effects, and of course specific period effects (eg school holidays). After accounting for all this, the residual day-to-day variations in demand will typically be classified as essentially random. We will also need to take account of random effects on the supply side, which will be predominantly due to “incidents” (including vehicle breakdowns, signal failures, burst mains etc.).

On the basis of this kind of reasoning, the Arup study (Arup 2003) adopted the definition of “JTV” [journey (or travel) time variability] as “unpredictable variation in journey times”, and went on to clarify:

One of the components of JTV is due to ‘incidents’, while what remains is referred to as ‘day-to-day variability’ (DTDV). and this in turn can be divided into two components: one due to unpredictable variations in demand, and the other due to random fluctuations in capacity, as represented in the following ‘equations’:

\[ JTV = DTDV + \text{Incident-related variability} \]
\[ DTDV = \text{Demand-related effects} + \text{Capacity-related effects} \]

Hence, in making any measurements of travel time variability, we need to define clearly what observations are in scope, and with respect to what the deviations will be measured.

Note that although this last sentence is cast in terms of what would typically be needed for the calculation of variance, measurements such as variance (and mean) are simply descriptive statistics for the distribution of travel time f(T) and don’t on their own define the full theoretical distribution.

1.4 Requirements for modelling and appraisal

In 1999, DfT’s ITEA Division reviewed the scope and aims of their research programme in respect of travel time variability, envisaging that ultimately “the generalised cost used in supply and demand modelling would take account of TTV just as it currently takes account of journey times, costs etc”. This implies that these two elements - demand modelling and the networks’ TTV performance (both highway and public transport) - need to be handled together, to ensure the two are in equilibrium, as illustrated in figure
1.2 below. It also implies that the third (appraisal) element of this process should be based on the origin to destination flows and TTV outputs of the combined demand and supply modelling process.

![Diagram of Equilibrium for generalised cost including TTV]

**Figure 1.2: Equilibrium for generalised cost including TTV**

Hence, this requires a) an agreed measure of travel time variability, b) a forecasting procedure (“supply model”) which can output the agreed measure in both a reference and “scheme” context, and c) a demand model which can respond to the difference in this output measure. Given all this, the supply and demand models need to be embedded in an equilibrium system (at least insofar as there are likely to be supply effects due to variations in demand), along the lines of current WebTAG recommendations for “variable demand modelling”.

As well as being potentially responsive to different levels of demand, the supply model is also the source of most policy input. Thus, in TTV terms, it will need to provide forecasts of TTV as a result of specific policies relating not only to standard network improvements and pricing but also (for example) the impact of improved signalling of rail junctions, provision of passing places for rail and buses, improved response to incident detection on the highway, etc.

It is essential to be clear about the different remits of the demand and supply models. As already indicated, while there is a reasonable amount of evidence relating to demand (and, related to that, **valuation**), the supply side modelling of travel time variability remains rudimentary. Of course, both sides must be in place if there is to be a workable overall model.
2. Theoretical methods for demand modelling and valuation of travel time variability

(Note: this chapter aspires to offer a reasonably succinct summary of the theory, some of which is discussed in more detail in Annex I).

2.1 Choice under Uncertainty

The fundamental proposition is that travel time is randomly distributed. The travel time distribution is, importantly, conditioned by the time of departure. For a given time of departure we may assume that in addition to the ‘free flow’ travel time there is also potential ‘recurrent delay’ due to congestion, such that the summation of the two is considered ‘predictable’. The first moment of this distribution accommodates both these notions, as well as the average of the purely random effects. The second moment of this distribution (the variance) provides one measure of the notion of ‘travel time variability’, distinct from recurrent delay in the sense that it is ‘uncertain’.

Given the general practice in transport modelling of making use of discrete choice theory, based on an underlying microeconomic “utility” theory, it will be useful to maintain this approach. However, the random nature of the travel time component means that that the utility itself becomes random, so that the discrete choice has to be exercised under conditions of uncertainty: we can refer to this as “Risk choice”.

As noted by Liu and Polak (2007) this notion of a stochastic utility is essentially different from the notion of “random utility models” in discrete choice theory, where the “error terms” reflect the modeller’s uncertainty about aspects of the choice process: here we are dealing with the case where, from the traveller’s point of view, some key aspects relating to the choice cannot be treated as deterministic. In random utility theory (RUT) we deal with situations in which there is a one-to-one correspondence between an act (e.g., implementing a travel decision) and the consequence of that act for the decision agent – i.e., it is a theory of riskless choice. In particular, the error term(s) in RUT accommodate inter alia unobserved heterogeneity and inter-alternative correlation but not agents’ uncertainty. Thus RUT does not provide an inherent treatment of risky choice – we need to extend it to do so. We can characterise this in Figure 2.1

In risky decision problems an act may give rise to several different consequences and we don’t know, ex ante, which will occur. We can therefore speak of choice between different prospects, with each prospect $r$ being defined as $(x_1, p_1; x_2, p_2; \ldots; x_n, p_n)$, where

---

2 The presentation and characterisation of risky choice problems and their relationship to existing travel demand models presented in this section draws extensively on the work of Liu and Polak (2007) and Michea and Polak (2006).

3 Note that while in general we refer to choice under uncertainty, in practice (from a microeconomic standpoint) we mean “risky choice”, which assumes that we know the probabilities of different possible outcomes. This is in some sense an approximation to true choice under uncertainty, which occurs when we don’t know the probabilities (which is of course the position).
each \( x \) represents the outcome vector of a possible consequence and each \( p \) represents the corresponding probability.

<table>
<thead>
<tr>
<th>Agent’s Uncertainty?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modeller’s Uncertainty?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Deterministic utility theory</td>
</tr>
<tr>
<td>Yes</td>
<td>Random utility theory</td>
</tr>
</tbody>
</table>

**Figure 2.1: Risky choice and Random Utility Theory**

This provides an entry point to the vast literature on economic choice under uncertainty. The dominant theoretical paradigm within this literature is expected utility maximisation (the classic approach is by von Neumann & Morgenstern (1947) - for a useful discussion of choice under uncertainty, see Deaton & Muellbauer (1980: §14)), though other variants have also been considered, as we discuss below.

According to this paradigm, choice outcomes follow a probability distribution, and expected utility is the expectation of utility across these choice outcomes. Given a risky prospect \( r = (x_1, p_1; x_2, p_2; \ldots, x_n, p_n) \) agents are assumed to seek to maximise

\[
V(r) = \sum_j p_j v(x_j)
\]

where \( v(x_i) \) is termed a (von Neumann and Morgenstern [vNM]) utility function, mapping the outcome vector \( x_i \) into utility space.

In other words, travellers choose the course of action which, bearing in mind the probabilities of different outcomes, has the highest expected utility. The approach implies that the traveller needs to assess all the eventualities resulting from different possible outcomes, though in practice, of course, he is likely to adopt a simpler version of this strategy.

A common variant on the expected utility model is to dispense with the use of the objective probabilities for different outcomes, i.e. those presented to respondents, instead
using ‘subjective’ probabilities. This allows for the possibility of respondents giving different weights to different outcomes. In this approach, the “true” probabilities \( p \) would be replaced by \( \pi(p) \), where this potentially also has a number of other parameters, and possibly also interacts with the attributes of the different outcomes. An in-depth discussion of this topic is given in Michea & Polak (2006), who test a number of different specifications, some of which outperform the standard expected utility model in terms of model fit. The outcomes of this study suggest that respondents tend to distort the presented probabilities, underweighting the probability of negative outcomes relative to positive ones, hence reflecting optimism.

Whereas the economic literature devotes its interest almost exclusively to uncertainty in monetary outcomes, this contrasts with our present interest in reliability, which embodies uncertainty in travel time rather than uncertainty in cost (acknowledging that the former may in itself incur some incidental cost). From the point of view of travel time variability, the most interesting choice to consider is that of departure time.

When the outcome vector contains a single variable (e.g., the amount of money gained or lost), then if the (vNM) utility function is linear in this variable, the expected utility will correspond identically with the utility in which the variable is set to its mean value. In such circumstances, the theory itself does not imply any loss of utility due to uncertainty. From this it can be seen that the agent’s attitudes to risk are embodied in the curvature (non-linearity) of \( \nu() \), as illustrated in the diagrams below. The so-called “risk premium” is defined as the difference between the mean value of the variable \( \bar{X} \) and the value \( \bar{X} \) at which the mean utility is obtained. In the first diagram, the risk premium is positive, demonstrating an aversion to risk (the standard case). By contrast, the second diagram depicts a “risk-prone” situation, where the agent is willing to pay to face a risky situation.
Risk premium $= \bar{x} - \hat{x} > 0$

Figure 2.2 Risk averse agent
Figure 2.3 Risk prone agent

However, in cases where the outcome vector consists of more than one variable, there is potential for confusion, as Liu & Polak (2007) have recently pointed out. Essentially, the concept of utility can be used in two subtly different ways. In line with RUT, utility can be thought of as a way of combining a set of variables into a single “metric”, usually, but not exclusively, using a linear form. An important example of this is the standard “generalised cost” metric combining cost and time variables. However, as we have seen, the vNM utility function carries with it some attitude to risk, and this is not present in the RUT formulation.

When the (RUT) formulation of utility is linear, so its’ expected value can be obtained by setting all random elements to their mean values, there is no reason for confusion: any “premium” associated with uncertainty must be dependent on the curvature of the vNM utility function. However, when the expected value of the (RUT) utility is not obtainable by merely substituting the mean value of the random elements, it is possible to generate an apparent uncertainty premium which does not require the introduction of a vNM utility function. We will see below that this is the case with the so-called “scheduling” model when it is used in conjunction with a travel time distribution.

In this latter case, Liu & Polak show that it is still possible to propose a von Neumann and Morgenstern utility function, and they demonstrate, using an appropriate dataset, that
an improvement in explanation can be obtained by doing so. On this basis, they propose that the term “expected utility” should be reserved for models where a vNM utility function is explicitly present, and the term “expected value” should be used for the case where we are merely taking the expectation of an RUT utility formulation, given randomness in some of the components, although it must be acknowledged that there is some disagreement within the team about the value of this terminology.

In practice, as noted by Liu & Polak, “the existing transport literature has combined the EUT and RUT perspectives in a variety of entirely plausible but nevertheless essentially ad hoc manners to draw conclusions regarding attitudes towards risk, principally, but not exclusively, those associated with variability in travel time.” The two major candidates in this respect are the so-called “mean-variance” approach, and the “schedule utility” approach, which we now go on to discuss.

2.2 The mean vs. variance approach

Suppose we write $U = U(T)$ where $U$ is a RUT utility function which we may assume is linear in $T$ (ie, in essence, of the generalised cost form), and $T$ is a random variable with distribution $f(T)$. In addition, we postulate a non-linear vNM function $v(U)$. It follows that the expected vNM utility is given by:

$$E[v(U)] = \int_0^\infty f(T).v(U(T))dT$$

Write $\mu = \int_0^\infty f(T).dT$ for the mean travel time. We can develop a general formula for $v(U(T))$ as a Taylor series about the value $v(U(\mu))$, ie:

$$v(U(T)) \approx v(U(\mu)) + (U(T) - U(\mu))v'(U(\mu)) + \frac{1}{2}(U(T) - U(\mu))^2v''(U(\mu)) + \ldots$$

Taking the expected value of both sides, this gives:

$$E[v(U(T))] \approx v(U(\mu)) + (E[U(T)] - U(\mu))v'(U(\mu)) + \frac{1}{2}E[(U(T) - U(\mu))^2]v''(U(\mu)) + \ldots$$

Now if $U$ is linear in $T$, then $E[U(T)] = U(\mu)$ so that the 2nd term on the RHS falls out, while $E[(U(T) - U(\mu))^2] = \lambda^2 \sigma^2$, where $\lambda$ is the coefficient on $T$ in the RUT utility function. Hence for linear $U$, we obtain:

$$E[v(U)] \approx v(U(\mu)) + \frac{1}{2} \lambda^2 \sigma^2 v''(U(\mu)) + \ldots$$

This shows that the expected utility can be approximated by a) inserting the mean value for $T$ in the utility function, and b) adding an additional term in the travel time variance.
(σ²), which is multiplied by half the second derivative of the Utility function at the mean. On this basis, we can write expected utility:

\[ E[\nu(U)] \approx \phi(\bar{T}) + k \cdot \sigma^2 \]  

(2.1)

Hence in these terms, the traveller will, in the face of travel time variability, choose the departure time that maximises his or her expected utility, where expected utility is a linearly additive function of the mean and variance of the travel time distribution. Empirical evidence suggests that both the mean and variance are “bads”, and it is implied that the term in σ² can be seen as a reflection of the “cost of unreliability”.

In practice, there is some ambivalence between this “pure” derivation using the travel time variance σ² (as used, for example by Jackson & Jucker (1981)) and approaches using the standard deviation of travel time σ, which has the advantage of being the same units as the mean. In a somewhat unrigorous way, both approaches are referred to as “mean-variance” approaches: most cases in the literature have, in fact, used the standard deviation. In addition, for empirical investigations, the implications of the functional form \( \phi(\bar{T}) \) have been ignored, and the most commonly applied function has the simplified form:

\[ E[\nu(U)] = \alpha \bar{T} + \lambda \sigma \]  

(2.2)

Practitioners commonly refer to the so-called ‘reliability ratio’ \( \tilde{\rho} = \frac{\lambda}{\alpha} \). This allows the relevant part of (2.4) to be written: \( \alpha \bar{T} \cdot (1 + \tilde{\rho} \cdot \frac{\sigma}{\bar{T}}) \), so that if the coefficient of variation \( (\frac{\sigma}{\bar{T}}) \) is constant, the allowance for reliability can be seen to be simply a “loading” for the value of time.

In practical terms within the transport field, models of risky choice based on mean and standard deviation have primarily been applied to route choice. Suppose a traveller has the choice of \( J \) routes, each of which has a travel time \( T_j \) which is a random variable with distribution \( f_j(T_j) \). A number of authors have suggested models in which the traveller chooses the route \( j \) according to the principle:

\[ \min_j [E(T_j) + \lambda SD(T_j)] \]

Essentially, this collapses the “prospect” for each route to an equivalent single alternative.
2.3 The scheduling approach

This derives from the pioneering work of Vickrey (1969) and Small (1982) to departure time choice, originally applied in a deterministic context. Later work by Noland and Small (1995) developed these ideas in the context of travel time variability, building on the earlier theoretical contributions of Gaver (1968) and Polak (1987).

One might see the scheduling approach as an attempt to instil the previous discussion with greater intuition in terms of travel behaviour. To this end, the scheduling approach reformulates the (RUT) utility function for a particular departure time choice as a function of four components; travel time, ‘schedule delay early’ (SDE), ‘schedule delay late’ (SDL), and a ‘lateness’ dummy variable that is set to unity if schedule delay late is non-zero. The latter three components are conditioned by the notion of a ‘Preferred Arrival Time’ (PAT), as follows.

- Journeys arriving before the PAT are deemed ‘early’. In this case, the SDE is derived as the difference between the PAT and the actual arrival time (i.e. the number of minutes of earliness), and SDL is zero.
- Journeys arriving after the PAT are deemed ‘late’. In this case, the SDL is derived as the difference between the actual arrival time and the PAT (i.e. the number of minutes of lateness), the lateness dummy variable is unity, and SDE is zero.

The four components of the utility function are typically specified as linearly additive, thus:

$$U = \alpha T + \beta SDE + \gamma SDL + \delta L$$  \hspace{1cm} (2.3)

where:

- $SDE$ is schedule delay early
- $SDL$ is schedule delay late
- $L$ is a dummy variable set to unity if $SDL > 0$, otherwise zero

Note, however, that because of the truncated form of the last three components, the overall function is not linear in travel time.

As in the case of the mean vs. variance approach, empirical evidence suggests that all components of utility are ‘bad’. If we now admit variability in travel time, then we can once again use the ideas of expected utility maximisation (but NB not assuming a vNM utility function), taking expectations of each component of utility over the travel time distribution.

$$E[U] = \alpha E[T] + \beta E[SDE] + \gamma E[SDL] + \delta E[L]$$  \hspace{1cm} (2.4)

where $E[L]$ may be re-interpreted as the probability of lateness.
Because of the discontinuous nature of the terms SDE, SDL and L, this expectation is not the same as would be obtained from inserting the mean value \( \bar{T} \) in the variable definitions: \( U \) is not linear in \( T \).

2.4 Equivalence between the mean vs. variance approach and the scheduling approach

Since the terms \( E[SDE] \) and \( E[SDL] \) embody variability in travel time, albeit conditioned by the notion of the PAT, it has long been suggested that there exists an (approximate) linear relationship between the summation \( \beta E[SDE] + \gamma E[SDL] \) and the standard deviation of the travel time distribution \( \sigma \). Strictly speaking, however, this approximation relies on departure time being continuously variable (as with the car mode, or with high-frequency public transport perhaps), and is closer for some distributions (e.g. exponential, uniform) than for others.

Until recently, a complete account of the conditions relating to this equivalence had not been demonstrated. Using the approach of Noland & Small, Polak (1995) had demonstrated it theoretically for the exponential distribution, when the optimum travel time is chosen. Appendix B of Arup Deliverable 6.1 (see Annex) takes this a little further, in also considering the uniform and logistic distributions, as well as one based on two “mass points”. It is noteworthy that the equivalence appears to be dependent on the form of the distribution.

In an important new paper, Fosgerau (2007) has demonstrated both the generality of the equivalence, and the dependence on the distribution of travel time. In what follows, we present his general argument.

For simplicity, we drop the “\( \delta \)” term and assume that the travel time distribution does not depend on the departure time\(^4\). \textbf{Whatever} the distribution \( f(T) \) is, we can represent the random variable \( T \) by a combination of the mean and variance, using \( Z \) as a standardized random variable with mean 0 and variance 1: in other words we write \( T = \mu + Z \sigma \) and write \( \phi(Z) \) as the distribution of the standardized variable. It can then be shown that the optimum departure time is given by

\[
\text{PAT} = (\mu + \sigma \Phi^{-1}(1 - \frac{\beta}{\beta + \gamma}))
\]

where \( \Phi^{-1} \) is the inverse cumulative distribution of \( Z \).

Substituting this into the utility function, it can be shown that we obtain an optimum utility equivalent to:

\(^4\) In Annex I it is shown generally how these restrictions can be relaxed
\[ \alpha \mu + \sigma \left\{ (\beta + \gamma) \int_{-\beta-\gamma}^{\beta} \Phi^{-1}(y) \, dy \right\} \]

Hence, at least under the restrictions stated, the reliability ratio \( \tilde{\rho} \) can be written as:

\[ \tilde{\rho} = \frac{\beta + \gamma}{\alpha} \int_{-\beta-\gamma}^{\beta} \Phi^{-1}(y) \, dy \]

This makes it clear that, for a given distribution \( \Phi \), holding the mean constant, the impact on the optimum value of the schedule delay terms in (2.3) is linear with changes in the standard deviation. For fixed values of \( \alpha, \beta, \) and \( \gamma \), the reliability ratio is fixed, but it is dependent on the distribution function \( \Phi \).

Fosgerau calculates the integral in the above formula, both for the normal distribution and for an empirical distribution which appears to fit the data for a particular road in Central Copenhagen. Using the commonly quoted Small (1982) coefficients for \( \alpha, \beta, \) and \( \gamma \) gives a value for the reliability ratio of 0.84 for the normal and somewhat higher (0.93) for the empirical distribution. As we will see, these are close to values which have been reported in the literature (e.g. 0.8).

It is noteworthy that the Arup work attempted some straightforward simulation work which changed the nature of the travel time distribution, and demonstrated that in some circumstances the implied benefits could be very different according to whether they were computed using the scheduling formulation or the mean vs. standard deviation formulation. This bears out Fosgerau’s conclusions about the dependence on the distribution.

It will be noted that the equivalence between the two approaches discussed here depends on the ability of travellers to continuously vary their departure time. This will in general not apply to public transport.

2.5 Applying the theory to the valuation of reliability

Typically we wish to apply the above theory to the (monetary) valuation of reliability. In order to do this, the convention is to supplement the expected utility functions (2.1) and

---

5 It should also be noted that while for some simple distributions, and for the normal distribution, the cumulative distribution of the standardized distribution is independent of the mean and the standard deviation, so that the reliability ratio is fixed, this is not generally the case. For example, with both the gamma and lognormal distributions, \( \Phi^{-1} \) is a function of the coefficient of variation (\( \sigma/\mu \)). Hence, for a given \( \mu \), variations in \( \sigma \) will lead to different values of \( \Phi^{-1} \). Hence the reliability ratio will not be constant. What is required is an understanding of how serious this is over the likely range of variation in \( \sigma \). This is not a criticism of Fosgerau’s mathematical treatment, but it means that the interpretation that can be placed on it is more limited than it at first appears.
(2.4) with travel costs, and to derive valuations of reliability as the marginal rates of substitution between cost and reliability variables. As already noted, in practice the mean vs. variance approach in fact usually considers standard deviation rather than variance (perhaps because standard deviation is measured in time units). If we expand (2.4) thus:

\[ E[U] = \alpha \bar{T} + \lambda \sigma + \theta C \]  

(2.5)

where \( C \) represents travel costs. (One might also wish to supplement with other variables depending on the practical context and policy interests, for example interchange and crowding.)

Then the marginal valuation of mean travel time (“value of time”) is given by \( \alpha / \theta \), and the marginal valuation of the standard deviation in travel time is given by \( \lambda / \theta \). The ratio of these valuations is, of course, the ‘reliability ratio’ \( \tilde{r} = \lambda / \alpha \).

Turning to the scheduling approach, we again supplement expected utility with travel costs as follows:

\[ E[U] = \alpha E[T] + \beta E[SDE] + \gamma E[SDL] + \delta E[L] + \theta C \]  

(2.6)

Then we can derive marginal valuations for each of the time and scheduling components, thus the marginal valuation of expected travel time is given by \( \alpha / \theta \), the marginal valuation of expected schedule delay early by \( \beta / \theta \), the marginal valuation of expected schedule delay late by \( \gamma / \theta \), and the marginal valuation of the probability of lateness by \( \delta / \theta \).

2.6 Challenges and opportunities for further theoretical research

Translating the theory to public transport

The theory as outlined above appeals particularly to the car mode, where it may be assumed that departure times are continuously variable. By contrast, public transport departures are often governed by fixed timetables, and this induces a discrete set of possible departure time choices. Bates et al. (2001) devote attention to this challenge, as does Batley (2007) from an alternative standpoint, but further work is probably needed. Such work would be usefully guided by an aspiration to harmonize the theory across modes, thereby permitting application to multi-modal analysis. Indeed there would be obvious attraction in proposing a single theoretical approach that is applicable to all modes.

There is evidence (eg Bates et al.) to suggest that in addition to scheduling costs with respect to PAT, public transport users may also take into account:
- Punctuality (i.e., compliance with the published schedule) which will in general be correlated with but not the same as the classical notion of schedule delay
- Pure variability (i.e., even if TTV does not make them late, they still may not like it)

Thus both mean-std and scheduling models may in fact be special cases of a general model.

*The interpretation of lateness, and alignment with industry conventions*

Following from the above point, and in particular the aspiration to harmonize theory, it would seem fundamental to seek greater consistency in how our notion of ‘lateness’ applies to different modes. In the empirical application of Bates et al. (2001), which relates to the rail sector, the scheduling model is supplemented with an additional variable representing mean delay (i.e. the operators’ notion of lateness, with reference to timetable). In many cases it appeared that the estimated coefficients of the lateness [mean delay] variable and $E[SDL]$ were not significantly different. This perhaps provokes the question of how public transport travellers interpret the PAT, and in particular whether they align the PAT with timetabled arrival times.

If the PAT and timetabled arrival time are one and the same, then this would seem to bring some convenience to the theory, as well negate the need to undertake the notoriously difficult activity of eliciting PATs from travellers. More generally, can the scheduling approach be re-couched without the need to explicitly specify the PAT? It might be noted that, in terms of modelling choice of time to travel, both PRISM and the Dutch national model operate in this vein, representing the time dimension in terms of large discrete periods (such as a two-hour peak period), and modelling shifts between periods as a result of changes in travel time and cost. Nonetheless, more thought would be required before it could be proposed for the investigation of reliability.

In order to illustrate the impact of unreliability in the public transport (in point of fact, rail) case, Bates *et al* made certain assumptions about the pattern of delays, and estimated the expected schedule disutility, conditional on the PAT. They concluded that, at least for low levels of variability, the impact of variability is dominated by the mean delay, especially for less frequent services (say, fewer than three trains per hour). This is in reasonable agreement with the existing practice of valuing unreliability for rail. Nonetheless, more generally the impact is sensitive to the distribution of delays, not merely the mean. This accords with intuition. In addition, they conclude that the actual distribution of PAT with respect to the timetabled arrival is potentially of great importance. Clearly, the impact on schedule delay of unreliability will be much greater when PATs are concentrated around the timetabled arrival. This is another area where more research is needed.

The work by Bates *et al* makes it clear that, with some effort, it is possible to apply a scheduling approach to the investigation of travel time variability on rail, but that the outcome is affected by a number of factors, not all of which are well understood. In spite
of the difficulties, it appears that the scheduling approach offers a greater chance of developing a harmonised approach across modes than does the mean-variance approach.

*Valuing travel time variability, as distinct from valuing time risk*

The distinction made earlier between the two types of utility function (RUT and vNM) and the possible confusion in some cases means that there is an unresolved debate as to whether the marginal valuations of reliability derived in the following section 3 encompass the full range of costs that unreliability imposes on the traveller. It should be noted that these valuations - whether derived from the mean vs. variance approach or from the scheduling approach - are strongly dependent on the form of the expected utility function. Acknowledging the earlier remark in section 2.1 concerning the analogy between uncertainties in money and travel time, it would seem instructive to compare marginal valuations of reliability with the conventions of the mainstream economic literature of choice under uncertainty.

Such a comparison provokes the suggestion that these marginal valuations, whilst adequately representing the impact of travel time variability on utility and choice, omit the cost of risk-bearing imposed on the traveller as a result of this variability, which may be referred to as the *risk premium*. If travellers are inclined to ‘insure’ against travel time risk, in the same manner that they routinely insure against money risk, then there is the potential for a risk premium.

Economic appraisal should, in principle, include any risk premia relevant to economic behaviour, and omission could potentially introduce bias. Further theoretical work is required to formalise the proposition of a risk premium for travel time, as well as empirical work to investigate its prevalence and magnitude (whilst this will likely require fresh data collection, it would be worth considering the potential for re-analyzing existing data).

This point is discussed at some length in Liu and Polak (2007) and also in Richard Batley’s recent work (Batley, 2007). The fundamental issue is the following. When we consider choice under uncertainty in the context of scheduling costs we are dealing with a non-linear system which, as demonstrated, induces a value of travel time variability. However, we can parameterise this system in different ways. In particular we can load explanatory responsibility either on the taste parameters that we are familiar with from choice contexts with **no** uncertainty, or we can add additional parameters that seek to characterise tastes with respect to risk, or we can do both. Each approach has different implication regarding what we mean by the value of variability. As noted, Liu and Polak have shown that models with an explicit characterisation of tastes with respect to risk outperform models without such a characterisation.

*Extension of choice dimensions*

According to the above theory, the only available response to unreliability is to adjust departure time. In the context of the present commission, it would seem pertinent to
consider whether a broadening of scope to include (at least) mode choice as well as departure time choice issues any challenge to the theory. There are examples from practice of joint mode choice-departure time choice models, but do they stand up to theoretical dissection? One might argue that mode choice can be straightforwardly included through the time, cost and scheduling variables. As we have already noted, however, some modes are constrained by timetabling whilst others are not, and this could reasonably feasibly introduce complications. Indeed, for any given traveller with a particular PAT, improvement in the reliability of one mode might yield benefit whilst improvement in the reliability of another mode may incur cost. Further complexities perhaps arise from multi-stage journeys involving various modes; in this regard Bates et al’s (2001) discussion of interchange offers a useful starting point.

Ultimately, of course, if we are successful in constructing a method for including travel time variability within some kind of “generalised cost” metric, then it will be necessary to address the same set of issues about choice hierarchy as are currently faced in constructing demand models. The discussion of mode choice was only singled out on the basis that a) some analysis of joint choice processes with departure time choice has already been carried out in this case, and b) there are different factors attending the departure time choice process when we are dealing with scheduled services, rather than continuous flexibility.

**Heterogeneity in valuations and PATs**

Finally, we should note that while the theory is typically expressed in terms of an individual traveller with a particular PAT, in practice it is necessary to take account of variations in PAT among the travelling public, as well as “tastes” relating to scheduling disutilities. We can expect that some segments will place a higher value on reliability than other segments, hence the call for methods that support a segmented analysis.

**2.7 Recommendations on theoretical approach**

Two versions of the theoretical paradigm to representing travel time variability have been presented here, the mean vs. variance approach, and the scheduling approach. In both cases, a strong recommendation is that further work is needed to progress the theoretical basis and to test the hypothesized model forms.

In our view, it is not currently possible to choose between these two versions without a clearer understanding of their similarities and differences. Certainly we concede that the mean vs. variance approach, together with the associated metric of the ‘reliability ratio’, has the advantage of greater simplicity, which gives it some appeal. This should not however detract from our belief that the scheduling approach may reveal important, and unique, insights into behavioural responses to unreliability, particularly in the context of scheduled (but infrequent) public transport services. Moreover, if theoretical methods are to permit meaningful comparison across modes, then any remaining challenges to the
correspondence between the mean vs. variance and scheduling approaches must be resolved.
3. Evidence on Valuation of Travel Time Variability

3.1 Introduction

This chapter deals with the question of finding monetary values for variability, discussing both passenger transport and freight, with more emphasis on the passenger side. Empirical outcomes are given in tables at the end of this chapter.

The key distinctions, which comprise the two main strands discussed in the theory chapter, used in this chapter are listed in Table 3.1.

| Table 3.1 Key distinctions in empirical studies on valuing travel time variability |
|---------------------------------|-----------------|-----------------|
| variation presented             | PAT obtained    | PAT not obtained|
| variation not presented         | C               | D               |

If in empirical studies variation in travel time is not presented and PAT (preferred arrival time) not obtained (D) then the data cannot be used for “reliability” purposes, so this case can be dropped.

However, scheduling parameters can be obtained whether or not variability is presented, as long as we have the PAT (or some reflection of it). We can then derive the response to variability by confronting the schedule delay formula with a distribution of travel time. These are cases A and C, which follow the schedule delay approach discussed before.

If PAT is not obtained, then clearly we have to resort to mean vs. variance approach (case B). We can also estimate mean/variance for case A.

As background to the work reported here and recommendations, however, it should be noted that practically all the empirical work that has been done to obtain values for variability, using either the mean vs. variance or the scheduling approach or both, has been based on Stated Preference (SP) data. It is generally very hard to collect Revealed Preference (RP) data that includes measures of variability, travel time and travel costs that will not be heavily correlated. Also, with RP data, there is the perennial difficulty of getting information on the attributes of the non-chosen alternatives, e.g. on the travel times at different moments (periods) in time. Nevertheless, researchers have tried to find situations with variability variation (e.g. the choice between two car routes where one is less reliable because of more congestion or bridges that might be closed) or between two train services. Examples of practical RP studies known to us are the ones carried out in California (SR91) comparing a route with a variable toll to an untolled (uncongested) route, and an ongoing study by ITS Leeds on rail time variability.

A few studies (e.g. Bates et al. (2001); Copley et al. (2002); Noland et al. (1998); Hollander (2005); Significance et al. (2007). . .) have presented both a measure of variation of travel time and obtained a PAT (case A in Table 3.1). Most empirical studies
have either used the mean vs. variance approach (case B) or the schedule delay approach (case C). The two dimensions noted in Table 3.1 both relate to significant difficulties for data collection. For cases A and B, there is the issue of presenting variability to respondents in a way which gives them sufficient assistance to provide reliable and useful data in terms of the choices offered. For cases A and C, a major difficulty is how to obtain the PATs. We discuss the difficulties relating to these two dimensions respectively in sections 3.2 and 3.3.

3.2 Presentational issues relating to variability (passengers)

In the context of travel time variability research, models often look jointly at valuations for the mean journey duration along with the valuation for the variance (or the standard deviation) of this journey duration. While this poses no issues as such from a modelling perspective, it is generally recognised that a non-trivial part of the population have difficulties in understanding the concept of the variance of journey duration in an SP survey. Aside from general design questions such as the number of alternatives and attributes, along with the attribute levels, the main issue that needs to be addressed in the survey design phase is the decision as to what approach should be used in the presentation of travel time variability in the survey questionnaires. A number of different possibilities arise.

The first possibility is to actually present respondents directly with a measure of travel time variability, such as for example the standard deviation in the travel time across a number of trips. The closest example of such an approach is the inclusion of travel time variability as an actual attribute in the surveys used by Hensher and colleagues in Australia. As an illustration, Figure 3.1 shows a screenshot from such a survey undertaken in Sydney, where this is taken from Hess et al. (2006).

As mentioned above, it is often argued that such direct measures of variability are difficult to understand for a non-trivial portion of the population. This is partly reflected in the low levels of significance obtained for such attributes (cf. Hess et al. (2006)), and also in the low relative valuations when compared to travel time (cf. Hensher, (2007)).

A very basic approach consists of presenting respondents with the probability of a certain journey being affected by issues with variability. An example of this approach is given in Figure 3.2, for the choice between two car routes in the context of the mobility pricing study undertaken by Vrtic et al. (2006). Here respondents face a choice between a tolled and an untolled alternative, where for the latter, in this case, one out of every 20 journeys takes at least 10 minutes longer than scheduled.
Figure 3.1: Example of survey including travel time variability as an attribute (taken from Hess et al. (2006))

Figure 3.2: Example of a survey design presenting respondents with the probability of a delayed journey (taken from Vrtic et al. (2006))
In the example shown, respondents are not given any information on the likely delay, although an indication of the lower limit on any delays is given (i.e. 10 minutes). In fact, studies often do not even give this information, and are based solely on the probability of a delay, or indeed the probability of arriving on time, such as in the example in Figure 3.3, taken from König (2004).

<table>
<thead>
<tr>
<th>Ich fahre mit</th>
<th>der Bahn</th>
<th>Dauer: 90 Minuten</th>
<th>Pünktlichkeit: 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ich fahre mit</td>
<td>dem PW die normale Strecke</td>
<td>Dauer: 45 Minuten</td>
<td>Pünktlichkeit: 50%</td>
</tr>
<tr>
<td>Ich fahre mit</td>
<td>dem PW den grossen Umweg</td>
<td>Dauer: 90 Minuten</td>
<td>Pünktlichkeit: 100%</td>
</tr>
</tbody>
</table>

**Figure 3.3: Example of survey providing respondents with a probability of arriving on time (taken from König (2004))**

In this example, the choice is between travelling with the train (“mit der Bahn”) with 100% certainty of punctuality (“Pünktlichkeit”), or by car (“PW”) either along the normal route with a 50% expectation of arriving by the stated time, or by a major deviation (“Umweg”) with a much longer time but with 100% certainty.

The approach of presenting respondents with a probability of a delayed journey can be extended straightforwardly by additionally providing them with a measure of the likely delay (when not zero), where this then again allows for a calculation of some form of travel time variability. An example of such an approach is given in Figure 3.4, again taken from König (2004). Here, respondents are presented with a choice between a slower and a faster alternative, where the latter (route B – a bus-rail connection) gets delayed by 30 minutes on 2 days per week (owing to the delays on the bus, which mean a missed connection). In this case, there is no further uncertainty as to the length of the delay, where such uncertainty can be incorporated by providing respondents with an anticipated mean delay in addition to the probability thereof, rather than a fixed delay if and when such a delay occurs.
Figure 3.4: Example of survey presenting respondents with probability of a delay and the extent of this delay

A different approach is to present respondents, for each alternative, with a set of possible travel times. This goes some way towards presenting the complete travel time distribution to the respondents, which from a research perspective would be the “ideal” representation of travel time variability. Often, this takes the form of presenting respondents with a set of possible journey durations (say 5 to 15) within each choice alternative, where this information is sometimes presented graphically. The researcher can then calculate the variance that is consistent with each set (or, more likely the other way around, generate a set of journey durations that matches a target variance). Average journey time and the variation in travel time presented in the SP survey can be chosen with a minimum level of correlation. Alternatively, respondents may be presented not with the actual travel times but with the early or late arrivals (compared to schedule) for a number of services. This is
for example the approach used in the work of Bates et al. (2001), where, as in many other studies of this type, a graphical approach (“clock-face”) is used to present the possible delays (cf. Figure 3.5). Here, respondents are told that each one of the possible outcomes has the same probability.

![Diagram of operator A and B patterns showing number of minutes early/late for typical ten train arrivals of London Paddington.]

**Figure 3.5: Example of a survey presenting respondents with a number of possible delays for each alternative (taken from Bates et al. (2001))**

The question invariably arises whether this method of presentation is not in fact more demanding for respondents than an approach giving a direct measure of travel time variability. As such, with the example in Figure 3.5, there is clearly a possibility that respondents do not in fact evaluate the alternative on the basis of mean delay and variation in the delay. Rather, they might simply work on the basis of the mean delay across the ten scenarios (e.g. 21.7 minutes late for operator A), or the probabilities of early and late arrival (e.g. 20% vs 80% for operator A).

In the Netherlands, a major empirical study commissioned by the Dutch Ministry of Transport is underway to measure the value to society of travel time benefits and travel time variability benefits in passenger and freight transport (see: Significance, VU University Amsterdam and John Bates, (2007)). The values will be estimated using Stated Preference experiments. In the past years several researchers have designed formats to present unreliability of travel times to respondents in SP experiments. These formats use concepts from statistics like “average travel time”, “travel time variance”, “probability of arriving late”, etc. To test whether these fairly advanced concepts are understandable for laymen travellers without a degree in statistics or even higher education, in-depth interviews among 30 respondents were carried out. The objectives of these (face-to-face) interviews were:

- Test the respondents’ understanding of different reliability presentation formats;
– Investigate the respondents’ assessments of these presentation formats with respect to clarity, ease of handling, and visual attractiveness;
– Collect the respondents’ preferences of the presentation formats.

In the analyses eight formats of presenting travel time variability were tested. All eight relied on giving five possible travel times within a single choice alternative, but differed in the wording and the use of graphics (e.g. bar charts, clock-face presentation as in Bates et al. 2001). Respondents were stratified according to their education level. There were three groups of questions. First, questions about how respondents conceptualise unreliability themselves. Do they think in terms of average, minimum, maximum travel time or probability? And how complicated do they find these concepts? Secondly, respondents were prompted with questions to test whether they gave the “right” answer for the different presentation formats. These test questions were designed to check to what extent the respondents have the “correct” perception of reliability, i.e., the same as expected by researchers, for each presentation format. Thirdly, there were questions about the respondents’ assessments of the eight presentation formats regarding clarity, ease of handling, and visual attractiveness, and which ones were preferred.

The interviews supplied a clear “winner” (see Figure 3.6) among the eight formats, which was the format without any graphical presentation, but just a list of five equi-probable travel times. This format is not only preferred by a majority of respondents, but also equally by people with low and high levels of education. What matters, however, is not what the respondents prefer but how well they understand the information that is presented. This format (with the corresponding arrival times added) will be used in the forthcoming main study.
Figure 3.6: SP presentation format that worked best in Dutch trial survey (taken from Significance et al. (2007))

3.3 Derivation of PAT for scheduling models (passengers)

There is a substantial body of work on the use of scheduling models (see the theory chapter), some of which yields empirical evidence on the value of travel time variability. The data used are often of an SP nature. A practical difficulty is that the model specification requires that the respondent elicits his or her preferred arrival time (PAT). It is by no means straightforward to ask this in an SP survey in a way that will correctly define what the researcher needs and will also be understood as such by the majority of respondents.

Some of the early scheduling studies (e.g. Small (1982)) have used the (fixed) start of the factory/office working hours as the PAT for commuting. Agreed starting or delivery times can also be used for (large sections of) business travel and freight transport, but usually do not exist for other travel purposes. Moreover, for commuting this approach has become less appropriate because of the increasing share of workers with flexible working hours.

An example of how questions into the PAT are phrased in a questionnaire survey (and how complicated things become) can be found in Significance et al. (2007):
1. The travel time of your trip that day was **TRAVEL TIME** minutes. How long would this travel time have been without any traffic jams or other delays?

   ____ minutes

2. Please keep thinking back to the trip you made that day. What would have been your preferred departure time if you had known **with absolute certainty** that there would be no traffic jams or other delays (so that the travel time would have been **ANSWER QUESTION 1** minutes)?

   ____ : ____  (use 24 hour notation, for example 15:45)

**CALCULATE PREFERRED ARRIVAL TIME FROM QUESTION 2 AND QUESTION 1**

3. This implies that your preferred arrival time would be **PREFERRED ARRIVAL TIME**. Is this correct?

   - yes
   - no, I would have liked to arrive at ____ : ____

Furthermore, the monetary values obtained for being early or late are very difficult to implement in a standard cost-benefit framework using the standard transport models, because the link to travel time period choice is not made in such analysis (there is no reference to clock time, only to journey durations), and the preferred arrival times are unknown. This however is a modelling and implementation issue (see next chapter), not a problem in valuation itself.

**3.4 Issues in valuation studies for freight transport**

The presentational issues are also relevant for freight transport. Just as in passenger transport, the concepts of variance and standard deviation are also often considered as too difficult for the respondents in a freight context, i.e. the shippers and carriers. Most studies use the probability of delay or the percentage of not on time arrivals instead.

In freight transport, the probability of delay is often measured as the probability of not arriving at the specified (by the shipper/receiver) time or within the specified time interval. This is then used as an equivalent of the PAT or preferred arrival time interval of passenger transport. The schedule delay could also include being too early, which leads to extra costs at the destination. Therefore, the outcomes are related to those of the scheduling approach. An explicit application of the scheduling approach to freight transport has also been undertaken by Small et al. (1999), and by Fowkes et al. (2001).
The values for variability in freight transport from the different studies are difficult to compare, because of the differences in the measurement units.

A specific difficulty in valuation studies for freight transport is who should be interviewed. Several agents are involved in decision-making on the same shipment (shippers, carriers, third party logistics service providers, truck drivers, ...). Massiani (2005) gives a theoretical argument that shippers will only give the VoT and travel time variability value of the cargo and carriers will include all elements (including costs for drivers, vehicles, cargo). We are not sure whether this argument will hold in a practical SP (which could also be seen as a game between shippers and carriers, with agents acting on how they would expect the others to act). Significance et al. (2007) makes the following assumptions (a priori hypotheses) on the freight VoT. Shippers with own account transport can give information on both the cargo-related and the vehicle/staff-related elements, shippers that contract out provide values for the first element and carriers for the second element. Of course there may be exceptions to the above general pattern, but in the freight questionnaires the shippers that contract out are steered (by very explicit instructions) to only answer on the components they generally know most about, and likewise for carriers.

There has also been some work in Australia on the valuation of variability in the area of freight, by Hensher et al. (2007), where they develop a framework that captures the interactive element of choice by incorporating ideas of concession and power for shippers and carriers, making use of mixture models (also see Puckett and Hensher, (2006); Puckett et al. (2007); Paglione et al. (2007)). In the survey, respondents are presented with the probability of an on-time arrival for the shipment, along with a host of other attributes.

3.5 Empirical findings for passenger and freight

It must be accepted the outset that the results which have been presented in individual studies are determined to a considerable extent not only by the way in which the data has been collected but also by the way in which it has been analysed. Therefore considerable care is needed in comparing the results of different studies. In almost all cases, a RUT approach has been taken to the analysis. With the exception of the recent work by Michea & Polak (2006) and Liu & Polak (2007), we are not aware of much work in this area which has explicitly specified a vNM utility function. A recent paper by de Palma and Picard (2005) did this, though not quite in the same way or in the same context. Other recent relevant work is by Recker et al (2005):

A key issue is whether the specification used in the models (in terms of the variables included in the utility function) is the same as that used in the presentation of the choice situations to respondents. It is of course important, in model specification, to stay as close as possible to the scenarios that respondents were actually faced with. In the case of data where respondents were presented, for example, with the probability of a delay, it is straightforward to include this directly in the same way in the models. This also still
applies when respondents are additionally faced with the average duration of any delay. As such, notwithstanding issues of respondent understanding, for the scenarios presented in Figures 3.1 to 3.4 it is possible to use the same specification in the models as in the survey data. The attributes relating to variability would be treated in the same way as the attributes relating for example to travel cost and travel time, typically using a linear-in-parameters approach.

The situation however becomes more difficult in the case of a design such as that used in Figure 3.5. Here, respondents are presented with a number of different outcomes, and it is clearly inappropriate to simply include the various outcomes in the utility function, each with their own estimated coefficient. Rather, we need to attempt to find a specification of the utility function that replicates the approach used by respondents in evaluating the various possible outcomes.

The number of possible approaches in this case is very large indeed, and the choice of an approach is difficult. There is a need to find an approach that attempts to do justice to the possible approaches used by respondents, but also one that is practical for use in modelling. The other issue is that by choosing a specific modelling approach this acts as a very strong assumption, imposing a certain evaluation tool on the data, where this may in fact differ significantly from the approach used by respondents. The other problem is that already at the stage of survey design, researchers might have an idea of the factors they are interested in, say for example the mean and standard deviation of travel time. They then produce a proxy for this information that they deem suitable for presentation to respondents and base their analysis of the observed choices on the original concept, say an evaluation of mean and standard deviation in travel time. When the assumed method however differs from that actually used by respondents, misleading results can be produced. This risk is potentially higher in the case of more complex specifications, such as percentiles or highly non-linear methods that may be difficult to translate into a simple set of equally likely outcomes. In short, the issue is to reconcile a mathematical specification of variability at the modelling end with the presentation of such information at the data end.

A summary of the main findings from the studies reviewed is presented in Tables 3.2 and 3.3. This includes quantitative outcomes, the method used in the study and a brief description of other relevant issues. Monetary outcomes from various empirical studies have been converted to Euros of 2003 (using international exchange rates and consumer price index numbers); some outcomes were in the original reports and papers only given in terms of minutes of travel time, and these were kept as such. It is clear from tables 3.2 and 3.3 that a very wide range of studies have been undertaken with outcomes which are not necessarily presented in a comparable format. Where comparable results are given there is still some diversity in the empirical values obtained. There are some apparent reasons for this, including the nature of the study, the sample taken and the level of disaggregation in the findings. This diversity in the evidence body is reflected in the overall recommendations outlined in section 3.5 below.
3.6 Recommendations on valuation

The review of empirical evidence on values for variability has highlighted the existence of a growing body of work on the modelling of the valuation of travel time variability. It should be acknowledged that this review is by no means complete. This is not helped by the fact that some of the existing work is not available in English, or is indeed not publicly available at all. As such, it can for example be assumed that there is a substantial body of work in the private sector, for clients such as railway companies, a point alluded to in König (2004).

While all reviewed studies agree that variability is a factor of substantial importance, there are no generally accepted monetary values for variability, or indeed a reliable estimate of the relative weight of travel time and travel time variability. It should be acknowledged that the valuations of variability in the present literature come from very specific investigations and are not even used in cost-benefit analyses in the respective countries of origin. Often, the studies are relatively small scale, and in some cases, variability is not the main topic of investigation.

The primary recommendation from this piece of work would thus be that there is a need for a major new study into the valuations of variability. Such a study would have to investigate both the question of how information on variability is presented to respondents and how it is later specified in econometric models that are estimated on the data. However, prior to this, agreement needs to be reached on the theoretical framework for journey time variability (as discussed in the previous chapter). The aim should be to achieve as high a level of consistency as possible between the underlying theory, the data collected and the estimated model, without unduly affecting understanding by respondents or the value of model outcomes. The a priori assumptions as to how respondents evaluate the information in the surveys should be kept to a minimum. Such work should also look further into the benefits of advanced modelling techniques such as discussed by Michea & Polak (2006) and Liu & Polak (2007).
Table 3.2. Value of variability (in travel time or Euros of 2003) in passenger transport: quantitative outcomes, methods used and other lessons.

<table>
<thead>
<tr>
<th>Study</th>
<th>Quantitative outcomes (+definition)</th>
<th>Method</th>
<th>Other lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent and HCG, (1995)</td>
<td>Doubling the chance of delay is equivalent to 13 min. travel time (commuting) or 20 min. (business and other travel); halving the chance of delay: 3 min. (commuting) or 5 min. (business and other travel).</td>
<td>Stated preference (SP) in road transport in the UK, with the following attributes: travel time, provision of information and chance of delay.</td>
<td>For some segments (e.g. business travel, time gains) the value of travel time per minute is higher in congested than in uncongested conditions.</td>
</tr>
<tr>
<td>ATOC, (2002)</td>
<td>Reliability ratio for train is in the range 0.6-1.5; most values in the range 0.8-1.3</td>
<td>Literature review as part of Passenger Demand Forecasting Handbook, UK</td>
<td></td>
</tr>
<tr>
<td>AVV, (2003)</td>
<td>Reliability (operating on time) is the most important aspect (importance of 3.58 on a scale from 1 to 5, with 5 being best) for bus, tram and metro, and the actual reliability performance is regarded as mediocre (5.94 on a scale from 1-10, with 10 being best).</td>
<td>SP among 3,387 users of bus, tram and metro in The Netherlands.</td>
<td></td>
</tr>
<tr>
<td>Bates et al. (2001)</td>
<td>Found significant valuations placed both on the inherent variability in travel time and (for scheduled services) on schedule compliance. Value of expected late schedule delay twice that of expected early schedule delay.</td>
<td>SP amongst 200 rail travellers</td>
<td>Underlying theory and recommendations for empirical research: scheduling model and SP data. Dealt with both unscheduled (car) and scheduled (public transport) contexts.</td>
</tr>
<tr>
<td>Brownstone and Small, (2002)</td>
<td>Value for 90th minus 50th percentile of the transport time distribution: 11-14 Euro/hr (males) and 28-30 Euro/hr (females).</td>
<td>RP: travel time measurements on State Route 91 in California, with variable tolls.</td>
<td>Method: can be done with RP data (in special cases such as this), use of percentiles.</td>
</tr>
<tr>
<td>Brownstone and Small, (2002)</td>
<td>Value for 80th minus 50th percentile of the transport time distribution: 26 Euro/hr.</td>
<td>RP (see above) and SP.</td>
<td>Travel time accounts for two-thirds of the service quality differential between the tolled and the alternative route; reliability one-third.</td>
</tr>
<tr>
<td>Copley et al. (2002)</td>
<td>The value of the standard deviation of travel time is 1.3 times the value of travel time (both per minute).</td>
<td>SP among 167 car drivers commuting in Manchester; mean versus variance method.</td>
<td></td>
</tr>
<tr>
<td>Copley et al, (2002)</td>
<td>1 minute late or early are valued less than 1 minute travel time.</td>
<td>SP (see above); scheduling model.</td>
<td>Method for valuation: mean versus variance approach or scheduling model.</td>
</tr>
<tr>
<td>Eliasson, (2004)</td>
<td>Reliability ratio of 0.95 for commuting 0.3 for business travel and 0.59 for other purposes (all car)</td>
<td>SP among 600 car drivers, Sweden</td>
<td></td>
</tr>
<tr>
<td>Hensher, (2007)</td>
<td>Value of travel time savings much higher than value of travel time variability</td>
<td>Travel time variability presented directly in SP.</td>
<td></td>
</tr>
<tr>
<td>Hollander, (2005)</td>
<td>Bus: mean travel time and early arrival valued at 8 Eurocent/minute; late arrival at 22 Eurocent/minute</td>
<td>SP among 244 bus users in York; scheduling model</td>
<td>mean-variance method gave no significant variance coefficient</td>
</tr>
<tr>
<td>Reference</td>
<td>Description</td>
<td>Method</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>De Jong et al. (2003)</td>
<td>Commuting, business and leisure travel: 1 minute late or early is 1-1.5 times as bad as 1 minute travel time; Education: 1 minute late or early is less important than 1 minute travel time; All purposes: 1 minute longer or shorter participation in activity at destination is less important than 1 minute travel time.</td>
<td>SP among around 1,000 car drivers and train users in the peak periods in The Netherlands; Scheduling model.</td>
<td></td>
</tr>
<tr>
<td>König, (2004)</td>
<td>Valuation of 34CHF for delay of 60 minutes, slightly lower for PT, 20% higher than value of time.</td>
<td>Various SP designs presenting direct measure of variability.</td>
<td>Important to look separately at probability of delay and length thereof.</td>
</tr>
<tr>
<td>Liu and Polak, (2007)</td>
<td>Value of risk aversion parameter is approximately -0.2, indicating a moderate level of risk aversion in the sample. Addition of risk aversion specification does not materially affect other model parameter valuations compared to Bates et al. results.</td>
<td>SP amongst 200 rail travellers (data from Bates et al.)</td>
<td>Proposes a (genuine) expected utility scheduling model with explicit characterisation of travellers’ attitudes towards risk. These model out-perform existing (expected value) models</td>
</tr>
<tr>
<td>Michea and Polak, (2006)</td>
<td></td>
<td>SP amongst 200 rail travellers (data from Bates et al.)</td>
<td>Proposes a number of non-expected utility models including various forms of reference point model. These models offer only moderate improvements compared to conventional models.</td>
</tr>
<tr>
<td>MVA, (1996)</td>
<td>Ratio of value of standard deviation of travel time to value of in vehicle time for car: business travel: 0.36, commuting and other: 0.78</td>
<td>Literature review, UK</td>
<td></td>
</tr>
<tr>
<td>MVA, (2000)</td>
<td>The value of the standard deviation of time in the bus is 24% of the value of travel time in the bus (when seated; less for travel time standing; both measured in minutes). The value of the standard deviation of waiting time is 48% of the value of waiting time.</td>
<td>SP among 309 bus users in France; Mean versus variance approach.</td>
<td></td>
</tr>
<tr>
<td>Rietveld et al. (2001)</td>
<td>A decrease in the probability of a 15 min. delay from 50% to 0% is worth 2.35 Euro (30% of the value of an hour travel time). A reduction in the probability of a 2 min.</td>
<td>SP among 781 public transport users in The Netherlands, with the</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Description</td>
<td>Methodology</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>delay from 50% to 0% is worth 0.32 Euro (therefore 1 min. delay is 2.4 times as bad as 1 min. travel time: risk-averse). Can be converted into a RR of 1.4</td>
<td>following attributes: travel time, probability of a delay, probability of a seat.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SACTRA, (1999)</td>
<td>By ignoring travel time variability the economic benefits of trunk road schemes are underestimated by 5-50% (UK).</td>
<td>Underlying theory: utility theory and attitude towards risk; Methods: SP and mean versus variance approach.</td>
<td></td>
</tr>
<tr>
<td>Senna, (1991)</td>
<td>The disutility of the standard deviation of travel time is around 2.5 times as high as for travel time.</td>
<td>SP survey among 301 respondents in Porto Alegre (Brasil), with a range of travel times, mean travel time and travel costs as attributes.</td>
<td></td>
</tr>
<tr>
<td>Stockholm public transport study, (2001)</td>
<td>Value of a minute delay is 3 times the VoT for metro and 4 for bus</td>
<td>Higher VoR for bus than for metro because higher share have connections that they may miss; indications of non-linearities: value of a 10 minute delay is less than twice the value of a 5 minute delay.</td>
<td></td>
</tr>
<tr>
<td>Swedish railway study (2004)</td>
<td>Value of a minute delay is 6 times the VoT or 100 euro/hour for private trips and 135 euro/hr for busbies trips</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>Vrtic et al, (2006)</td>
<td>Reduction in proportion of late trips by 1% valued at 0.5CHF for car travel, and 0.2CHF for public transport.</td>
<td>Probability of late arrival included as attribute in SP.</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3. Value of variability in freight transport (in travel time or Euros of 2003): quantitative outcomes, methods used and other lessons.

<table>
<thead>
<tr>
<th>Study</th>
<th>Quantitative outcomes (+definition)</th>
<th>Method</th>
<th>Other lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent and HCG, (1995)</td>
<td>A 1% increase in the probability of delay of 30 or more min. is equivalent to 0.45 – 1.8 Euro of 2003 per transport.</td>
<td>Stated preference (SP) in road transport in the UK with the following attributes: travel time, travel costs, provision of information and probability of delay. Method: SP</td>
<td></td>
</tr>
<tr>
<td>Bogers and van Zuylen, (2005)</td>
<td>Truck drivers value the unfavourable travel time twice as high as its objective (risk-neutral) worth. Managers of shippers and carriers did not have this relatively higher value for unfavourable travel times. This measure of unreliability cannot easily be transformed into a reliability ratio for the standard deviation of travel time.</td>
<td>SP among truck drivers and managers of shippers and carriers, used a visual presentation with one favourable travel time once in 10 days, one unfavourable and 8 normal travel times</td>
<td></td>
</tr>
<tr>
<td>Bruzelius, (2001), based on Transek, (1990), (1992)</td>
<td>Sweden: for rail transport, a 1% increase in the frequency of delays is equivalent to 4.7-7.0 Euro per wagon; For road transport: 3.5-32.6 Euro per transport.</td>
<td>SP survey among shippers in Sweden in 1989/1990, including the following attributes: costs, transport time and probability of delay.</td>
<td></td>
</tr>
<tr>
<td>Fowkes et al. (2001)</td>
<td>UK, road transport: the value of the difference between the earliest arrival time and the departure time is on average 1.18 Euro per min. per transport (more or less the free-flow time); for the time within which 98% of the deliveries takes place minus the earliest arrival time, the value is 1.44 Euro (‘spread’); for deviations from the departure time (schedule delay) the value is 1.12 Euro.</td>
<td>SP survey among 40 shippers and carriers in the UK in 1999 with the following attributes: time, costs, latest departure time, earliest arrival time, arrival time for 90, 95 and 98%. Method: SP</td>
<td></td>
</tr>
<tr>
<td>Hensher et al. (2007)</td>
<td>Valuation of reliability gains of 2.20 Euro per percentage point for transporters, 6.50 Euro for shippers. This is obtained when looking solely at the freight rate; when further incorporating all costs in the calculation, the VRG rises to 7.90 Euro. Giving an actual meaning to these values, the results would imply that, if a toll free route had a 91% probability of on-time delivery, with 97% for the tolled route, the value of trip time variability for transporters would be 13.30 Euro per trip.</td>
<td>Mixture models incorporating ideas of concession and power, probability of on-time arrival presented in SP.</td>
<td></td>
</tr>
<tr>
<td>HCG, (1992a)</td>
<td>The Netherlands: an increase in the percentage not on time by 10% (e.g. from 10% to 11%) is just as bad as 5-8% higher transport costs.</td>
<td>SP survey in 1991/1992 among 119 shippers and carriers in goods transport by road, rail and inland waterways with the following attributes: time, costs, Method: SP.</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Context</td>
<td>Method</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>HCG, (1992b)</td>
<td>A decrease in the probability of delay by 10 index points (e.g. from 15% to 5%) is worth 0.5 – 2 Eurocent per tonne-km.</td>
<td>SP surveys in 1992 in The Netherlands, Germany and France with around 50 interviews per country with the following attributes: time, costs, probability of delay, frequency and flexibility.</td>
<td>Method: SP.</td>
</tr>
<tr>
<td>MVA (1996)</td>
<td>Reliability ratio for transport: 1.2</td>
<td>Literature review</td>
<td></td>
</tr>
<tr>
<td>RAND Europe et al. (2004)</td>
<td>The Netherlands: a change of 10% in the percentage not on time (e.g. from 10% to 11%) is equivalent to 1.77 Euro per transport for goods transport by road. Also values for rail, inland waterways, sea and air transport.</td>
<td>SP/RP survey among 194 shippers and carriers in road transport with the following attributes: time, costs, percentage not on time, probability of damage and frequency.</td>
<td>Method: SP model.</td>
</tr>
<tr>
<td>Small et al. (1999)</td>
<td>USA: A reduction in the deviation from the agreed delivery time (schedule delay) by 1 hour is worth $393 Euro per transport.</td>
<td>SP survey among hauliers in the USA; Scheduling model.</td>
<td></td>
</tr>
</tbody>
</table>
4. Supply issues

As was made clear in Section 1.4, there has been a good deal of work in relation to modelling TTV in the demand area. However, if TTV is to be brought into the general transport modelling domain, it is essential to also represent the supply effects, and in this area there has also been a great deal of research. In this Chapter we discuss the issues, with examples of the current state of theory and practice. Although the work has addressed the performance of both public and private transport networks, it is probably fair to say that the majority of the work has focused on the highway context. This balance between highway and public transport network modelling is also reflected in UK planning practice, where the tools applied for highway modelling are much further advanced than those typically applied for public transport modelling, and goes beyond the question of whether and how to model TTV. Because of these two considerations, our review therefore focuses on the highway context, before considering briefly the position for public transport. A further point to make about the scope of this chapter is that we will not consider micro-simulation methods. Although these are becoming increasingly used in the UK, the route-choice/demand side of micro-simulation tools is relatively poorly developed, at least in practice, meaning that their application has so far almost exclusively focused on supply-side issues. The starting point of this Chapter, by contrast, is that a combined approach is needed, which deals consistently with demand, route choice and supply-side factors.

The Chapter begins by briefly describing the basic principles of state-of-practice highway assignment modelling, without TTV. We then move on to consider the issues for introducing TTV within such approaches. In order to bring some structure to the discussion, we divide the modelling discussion into two broad issues, namely the largely empirical issue of how the actual variation in travel times is represented (section 4.2), and the largely theoretical/algorithmic issue of how network performance is modelled in the light of TTV (section 4.3). From this review, we see that many developments are taking place in the first of these areas that give cause for optimism regarding practical implementation, whereas the second area has some practical constraints (identified in detail in section 4.4) that do not appear to be the focus of any present study. Based on the constraints identified, we move on to consider how to take forward the modelling of network performance under TTV (section 4.5). With the focus of the chapter on highway network assignment, we briefly consider in section 4.6 the distinct issues that arise in public transport networks. In section 4.7 we conclude with a summary and recommendations for how to take this work forward.

6A point of terminology is that in this report we shall refer to the role of the network model as a ‘supply-side process’, although this is not a universal usage of this term. In particular, the route-choice process could be considered to be a component of demand, and in that case it is the implementation of capacity restraint on the links of the network that represents the true supply function; this is the terminology that would be more commonly understood in network and/or traffic modelling. However, as the current report has been written at the outset from the perspective of demand modelling, we have found it more natural to adopt the terminology that is common to that field, where the understanding of the role of an assignment model is as a method for managing the interface between demand (essentially between pairs of zones) and supply (essentially at the network/link level).
4.1 Highway assignment models: state-of-practice and issues for representing TTV

Network models are one of the commonly used tools to forecast changes in flows, demands and travel times under the conditions of a newly applied policy instrument. Typically such models assume driver decisions are motivated by the generalised cost, composed of a combination of distance (as a proxy for vehicle operating cost), travel time and any tolls payable. In this context, we might ask: what exactly do we mean by ‘travel time’? In normal circumstances, it is presumed to represent an average, but if we also wish to reflect that travel times are variable, then how might we achieve this with such models?

Since it will be important for the subsequent argument, it is worth spending a little time discussing this standard procedure, and introducing some notation. The aim is not to give a complete account of highway assignment, but to set down the main points.

Let us consider the simplest form of highway assignment model that represent well the current state-of-practice: fixed demand, one user class, static, deterministic (Wardrop) user equilibrium. The historical origin of the name ‘assignment’ in transport derives from the first all-or-nothing style models, in which the purpose was to take a matrix of vehicle travel (as movements from origins to destinations) and to assign it onto an appropriate network. Subsequently equilibrium models extended this notion to the representation of a kind of ‘game’ in which the aim is to find a consistent solution to several processes, which may be described as:

1) drivers on each origin-destination movement choosing minimum cost paths through the network, with cost including travel time as an attribute;
2) the flows on the links of the network being an aggregation of the flows on all paths for all origin-destination movements;
3) dealing with supply-side effects (capacity restraint) as a result of the volume of link flows relative to capacity, implying that the (mean) link travel times are a function of the link flows (possibly including junction interactions).

It has long been known that such fixed demand approaches may be extended to handle at least some forms of demand response; alternatively it may be desirable to model the demand responses by some external model, in which case one may also consider the output of such a model to be the travel costs on each origin-destination movement. Note that such a concept is well-defined, at least for the deterministic user equilibrium model, since at equilibrium all the used routes for a given OD movement have the same travel cost: so there is a unique notion of an OD travel cost. (The same is not true for stochastic user equilibrium models, or models where the outputs themselves are stochastic, where different measures need to be considered).

In line with the terminology of Clark & Watling (2005), we will define:
\( v_a \)  flow on link \( a \) (\( a = 1, 2, \ldots, A \)),
\( \mathbf{v} \)  the vector of flows across all links
\( q_w \)  mean demand on O–D movement \( w \) (\( w = 1, 2, \ldots, W \))
\( \mathbf{q} \)  \( W \)-vector of mean demands
\( R_w \)  index set of acyclic paths serving O–D movement \( w \)
\( f_r \)  flow on path \( r \) in \( R_w \) serving O–D movement \( w \)
\( \mathbf{f} \)  vector of path flows across all paths and O-D movements
\( \delta_{ar} \)  indicator variable, equal to 1 if path \( r \) contains link \( a \), 0 otherwise
\( t_\alpha(\mathbf{v}_a) \)  (mean) travel time on link \( a \) as a function of \( \mathbf{v}_a \) (\( a = 1, 2, \ldots, A \))
\( \mathbf{t}(\mathbf{v}) \)  vector of functions \( t_\alpha(\mathbf{v}_a) \) (\( a = 1, 2, \ldots, A \))

Then a fixed demand Wardrop User Equilibrium (UE) model with travel time the only component of travel cost, is defined as an assignment of flows and costs to the network that simultaneously satisfies the following conditions (with an asterisk denoting equilibrium values of the variables concerned):

1) Equilibrium route costs are consistent with link costs that would arise at the equilibrium link flow levels:

\[
\mathbf{c}_r^* = \sum_{a=1}^{A} \delta_{ar} \mathbf{t}_{\alpha}(\mathbf{v}_a) \quad (r \in R_w; w = 1, 2, \ldots, W). 
\]

2) Equilibrium link flows are consistent with the equilibrium path flows:

\[
\mathbf{v}_a^* = \sum_{w=1}^{W} \sum_{r \in R_w} \delta_{ar} \mathbf{f}_r^* \quad (a = 1, 2, \ldots, A). 
\]

3) Equilibrium paths flows are consistent with the OD demands:

\[
\sum_{r \in R_w} \mathbf{f}_r^* = \mathbf{q}_w \quad (w = 1, 2, \ldots, W). 
\]

4) Wardrop condition is satisfied for each OD movement: used routes having equal cost (which we can therefore term the OD cost), and routes with higher cost carrying no flow:

a. For each \( w = 1, 2, \ldots, W \), then \( \mathbf{f}_r^* > 0 \Rightarrow \mathbf{c}_r^* = \pi_w^* \) and \( \mathbf{c}_r^* > \pi_w^* \Rightarrow \mathbf{f}_r^* = 0 \) \((\forall r \in R_w)\)

b. where \( \pi_w^* \) is therefore the equilibrium OD cost for movement \( w \).

Some care needs to be taken in interpreting this definition: for example, it is well known that the equilibrium path flows are non-unique, and so typically we envisage the output of such a model being in terms of entities that can be guaranteed to be unique (under certain conditions), namely the link flows, and link/path/OD costs. For a Stochastic User Equilibrium model a somewhat different form of definition is possible, since in such cases there is unique notion of the route flows (and the route choice proportions). To maintain maximum similarity with the definition above for UE, it is probably most convenient to imagine the SUE model requiring the same conditions 1-3, but replacing condition 4 with the SUE requirement:

For each \( w = 1, 2, \ldots, W \), then \( \mathbf{f}_r^* = q_w \mathbf{p}_r^*(\mathbf{c}^*) \)
where for any vector of route costs \( \mathbf{e} \), the function \( p^* \mathbf{w} (\mathbf{e}) \) denotes the proportion of demand that would use route \( r \in R_w \) for movement \( w \). Note that a similar form of definition is not possible for the UE model as such a well-defined function \( p^* \mathbf{w} (\mathbf{e}) \) does not exist in the UE case.

Before proceeding we should note that a further aspect which needs to be discussed is the consequential impact of the supply model output on demand. In terms of developing an understanding of new techniques, it makes sense to focus initially on the simpler inelastic demand case, where the only response represented in the model is one of route choice. However, it will also be appreciated that standard practice in network assessment is now to perform a variable demand evaluation, in the spirit of the VADMA advice and in conformity with DfT guidance. Thus, whatever approach is developed will ultimately require to be embedded within a supply-demand equilibrium system. While this may be viewed as primarily an algorithmic issue, in terms of whether the convergence of such a system can be guaranteed/attained at all and the run-times that would arise, it is also likely to have practical implications in terms of the generality of models that may be accommodated, and the consistency of the theoretical assumptions made about perceptions of TTV on the demand-side and the route-choice side. All such issues will need to be borne in mind at some stage; we begin to address some of them in subsequent sections, even though our primary focus is on the inelastic demand case.

Having set out the preliminaries, therefore, for conventional, fixed demand assignment modelling, the question is then: how does TVV fit into this framework? This question has several facets to it. In order to bring some structure to the subsequent discussion, we divide the modelling discussion into two broad issues that any approach to modelling TTV must consider:

1) How is the actual variation in travel times represented and modelled?
2) How is network performance affected by TTV, incorporating both the behavioural (route choice) response to TTV and the impacts this has on the equilibration process.

The first issue, of representing actual TTV, is considered in section 4.2, and is seen to divide once more into two further classifications, namely those approaches that aim to model the composite effect of all sources of TTV (without regard to the causes), and those approaches that aim to represent the distinct impacts of the various causal factors, notably demand variability and incidents.

The second issue, of modelling network performance under TTV, is a much wider issue, meaning that it is not feasible to review all the theoretical approaches that have been proposed for dealing with it. Instead in section 4.3, we choose to illustrate some of the key considerations that arise in modelling network performance by reference to a sample of approaches in the literature. Throughout we pay particular attention to the relevance of previous DfT-supported research in these areas. Although this broad, two-way classification is not ideal, as there are clearly overlaps between the decision as to how to model actual TTV and the decision as to how to model network performance under TTV,
it at least gives some way of structuring a complex research area, and some way of comparing the disparate approaches followed.

### 4.2 Representing actual TTV in highway assignment models

As discussed in section 4.1, an issue that cuts across the question of the overall framework to be used for network performance under TTV is: how do we actually represent the TTV itself (rather than the drivers’ or system’s response to it)? Clearly TTV is not constant across the links and routes of a network, and is certainly likely to be policy-sensitive. This suggest that as a minimum (and for ease of connection to observable data on variability), we would consider specifying variability on a link level. For example, in this context we could mention the work by Arup (discussed previously in section 1.4), who showed that with a linear speed-flow relationship of the form $\alpha_a - \beta_a v_a$, where $v_a$ is link flow and $\alpha_a$ is the free-flow speed, it is possible to derive a formula for the coefficient of variation of link travel time on link $a$, $CV_a$:

$$CV_a = \sqrt{\text{var} \alpha_a + \beta_a^2 \text{var} v_a} / (\alpha_a - \beta_a v_a)$$

and that this can be generalised, without substantial effort, to more complex speed-flow curves.

Taking a somewhat different approach, Arup (2003) have also developed original work by SDG (1993), based on the observation that TTV is likely to be greater as flows reach capacity. This observation motivated the use of the so-called congestion index CI as a key explanatory variable in representing TTV, the CI being defined as the ratio of the mean travel time to the free flow travel time for a journey (note: not for a link). In particular, it has been proposed to relate the coefficient of variation ($CV = \text{ratio of standard deviation to mean}$) of travel times to the CI, which gives model-users a simple way of generating travel time variances from standard data that already exists on mean travel times. Using data collected from routes in London and Leeds, relationships between CV and CI were estimated, and are reported in Arup, 2003 (chapter 15). In carrying out the estimation, it was noticed that there was a tendency for the journey CV to decline with journey length $d$. The results for the two datasets were not identical, but showed some degree of consistency, and therefore in order to achieve some kind of transferability to other locations, the consultants proposed a “compromise equation” giving a relative weight of 75:25 in favour of the Leeds results, which was of the form:

$$CV = 0.148 \ CI^{0.781} d^{-0.285}$$

and this has been proposed in the latest WebTAG guidance (Unit 3.5.7) for use in urban studies. An important point to note about this work is that it is based on relationships between CV and CI for a whole journey, not for individual links of a network: the relationship proposed does not decompose in a way that we can consistently assume it to hold on both a link level and a journey level.
There are several other, on-going works of relevance in this area, whose results when reported should further enrich such modelling. On-going DfT-funded research by Hyder has the aim of “develop[ing] sound theoretical models supported by robust data to give firm underpinning to future estimates of the relationship between travel time variability traffic flows and journey time.” In a similar spirit, on-going work by Mott MacDonald for the DfT (yet to be reported) has used extensive inter-urban data from the HATRIS database to derive various forms of model, relating travel time variability to link (rather than whole trip) attributes, for different link types and day-types, thus providing potential for link-based variability inputs to be readily generated for network models. Alternative model forms for inter-urban studies are likely to arise from this research.

However, a somewhat different approach to this problem is to start with a belief that we need to explicitly model the *causes* of TTV, rather than seeing TTV as either a fixed, inevitable element or one that is only linked to the mean link travel time. In this respect, we might refer to the Arup work noted in Section 1.4, which adopted an approach in which highway TTV derives from two general sources, what they referred to as “incident-induced” variability and “day-to-day variability” [DTDV]. The latter term refers to the variability in travel time brought about by changes in demand from one day to the next, considered at the same time of day, and allowing for seasonal and other, essentially predictable, effects. This consideration implies that the earlier, CV-related work described above describes the composite effect of variability from both DTDV and incidents. An argument in favour of the (composite) CV approach is that a) no evidence has so far been adduced that the demand response is different between the two sources, and b) at least in an urban context, the impact of most incidents is likely to be difficult to distinguish from that of the random demand fluctuations. The counter argument is that by representing the underlying causal factors, we may gain a deeper understanding of TTV and introduce more potential policy levers to affect TTV in our models. Adopting this kind of philosophy, allowing for DTDV would in principle evaluate the impact on link travel time variability by allowing explicitly for the variation in the OD demand matrix **q**. We discuss this further in the next section.

Continuing with the line of modelling the underlying causes of TTV, we should mention that the treatment of incidents is conceptually separate and rather more problematic than that of representing DTDV. These are essentially capacity-reducing random events whose frequency is nonetheless sensitive to the level of demand, and whose consequences are also sensitive to the level of demand, in terms of the length of queues that may be generated. One approach to conceptualising such incidents is as follows: Define an incident as a discrete event E with random characteristics **h** (which may include the type of incident, the amount of the carriageway blocked etc.). On any given link, the random probability of E occurring is conditioned by the flow v: \( p[E|v] \). Given E, the consequent delays are a function of **h** and the flow v. Once again, therefore, if we could input the associated probabilities (essentially those of an incident of given characteristics occurring), then we could generate the outcome distribution in terms of travel time. It will be appreciated, however, that this is an even more demanding task than the case of DTDV.
The impact of incidents is likely to be greatest when alternative routes are limited, as in the case of Motorways. It is precisely in such cases that explicit attention may be given to projects for reducing the effects of TTV due to incidents. The interest in this topic has led to the work underlying INCA (which in turn derives from INCIBEN), which has been funded by the Department. This was carefully reviewed within the Arup work, together with a comparable approach developed in the US by Cohen & Southworth (1999). Chapters 5 and 6 of D6.1 (Arup, (2002)) provide a clear theoretical account. Key factors noted by Arup (Chapter 2) are as follows:

- The impact of an incident is different from the DTDV case in that the incident occurs at a random time (and place), and this random time of occurrence itself induces some variability in terms of impact.
- If the motorist arrives before the incident occurs he will experience no delay. If he arrives after the queue has dispersed he will again experience no delay. Otherwise, the delay that he experiences depends on his time of arrival relative to the time of the incident. If the total duration of the incident is $T$ minutes, then the probability of a motorist experiencing some delay is equivalent to the probability of the incident occurring up to $T$ minutes earlier than the time of “arrival”, while they argue that the delay experienced by an individual motorist who encounters a queue can be considered to be uniformly distributed.
- This allows us to calculate the overall mean and variance of delay from a randomly occurring incident, taking into account the probability of occurrence. An important result is that the mean is proportional to the square of the incident duration, while the variance is related to the third and fourth powers.

Arup considered how to allow for random characteristics of an incident, in terms of its duration and severity: by assuming either a theoretical distribution, or an empirical distribution drawn from a sample of observations based on recorded data from actual incidents, as is done in the INCA methodology. The conclusions of their analysis was that it is legitimate, as an approximation, to add the delay variances over incident types and links, as is done in INCA, provided: (a) the square of expected delay is small relative to expected value of the squared delay; and (b) not more than one incident (causing delay) is encountered during the course of the journey. The underlying basis is currently being extended to single carriageways, it being originally for motorways only and later extended to grade-separated dual carriageways. The program is recommended in WebTAG Unit 3.5.7. Although the previous version of INCA did not incorporate an explicit calculation of TTV due to DTDV, the new INCA 4.0 now models DTDV explicitly.

In summary, there is a great deal of work in the area of representing actual TTV, whether in composite form or by its component causes, much of this work DfT-funded, and several studies on-going at present. The understanding emerging from these studies is sufficient to believe that specifying models of TTV is feasible in realistic networks, and such models will only improve with further investigation of empirical sources.
4.3 Modelling highway network performance under TTV

As explained in section 4.1, we have divided the modelling of highway networks under TTV into two broad themes; in the previous section we considered how we might represent the actual variation in travel times, without regard to how the drivers or the network may be impacted by this variation. In the present section, on the other hand, we presume that suitable models of actual TTV exist (whether composite or as separate cause-related models), and the key question is how to model network performance in such an environment. In other words, if travel times are variable, how do we take forward the conventional UE/SUE approaches described in section 4.1?

Taking a software-oriented viewpoint, a common first response to this question runs as follows. Let us consider, purely for illustration, a cause-based approach to representing TTV based on representing DTDV effects only; that is to say, stochastic travel times arise from the stochastic demand. It seems reasonable, then, that we could proceed along the following lines, using Monte-Carlo (MC) methods:

Assume a probability distribution for the day-to-day distribution of demand \( q \)

1. Simulate day \( n \), initially with \( n=1 \):
   - Randomly select a demand \( q^{(n)} \) from the assumed demand distribution.
   - Assign \( q^{(n)} \) using a standard UE/SUE approach.
   - Record the link travel times \( t^{(n)} \) that arise from assigning this demand.
   - Increase \( n \) by 1 and return to simulate the next day, until \( n=N \).

Collect together the results from all \( N \) days to give an empirical distribution.

Conceptually the ideas seem straightforward, and we are left only with a numerical challenge of replicating many equilibrium assignments. In fact, we may circumvent this numerical speed issue by using an approximation to the equilibrium model, known as sensitivity analysis, which means that we can produce the same (in fact better) results as the process above in a fraction of the time, using a single equilibrium assignment run. Clark & Watling (2005) critique this kind of approach in some detail, which has been suggested in several previous studies, and highlight the key flaw in its reasoning. Namely, since equilibrium is reached on each day, it implies that drivers have perfect predictive knowledge of the travel times they will experience on that day, before they make their journey. This seems a reasonable assumption for slowly-changing or systematic trends where drivers may have an opportunity for repeated experience: for example, if the different samples were to represent the different mean demands arising in school-holidays and term-times. For events that can change on a daily basis, such as demand and incidents, it seems more difficult to justify however, and so even though the
resulting problem is tractable it is highly questionable as to whether it is the appropriate approach for representing the kind of TTV impacts we have in mind here.\footnote{Just to emphasise the fact that the reason for rejecting the Monte Carlo procedure proposed at the start of the section is one of appropriateness, rather than its computational difficulty, it does seem that this kind of approach is appropriate for other areas of the Department’s interests, such as representing the impact of uncertainty in traffic forecasts due to limitations in the input data provided to a model. In a different source paper to that cited above, Clark & Watling (2006) describe the use of such an approach (comparing Monte Carlo and more efficient analytic methods), in which the probability distribution of demand now represents the sampling error of the observer/planner in estimating the current mean demands, and the objective is to derive a confidence interval for the resulting model outputs that reflects this uncertainty in the input data.}

Although the process described above does not, then, seem the appropriate solution, it is useful as a reference point to consider how we might go about incorporating TTV in a suitable way. In particular, it is clear that a representation of TTV in the model is an acknowledgement of some kind of uncertainty – but whose uncertainty? Once we are clear on this question, we are much closer to finding an appropriate method. Two alternative answers to this question that will motivate the work presented below are:

a) The drivers have uncertainty in making their route choice decision.

b) The planner has uncertainty, even if the input data to the model are error-free, because the actual day-to-day performance of the network is variable daily.

A rather fundamental question to consider first is: does TTV bring into question our whole equilibrium framework? Since the equilibrium models, be they UE or SUE, all implicitly assume that travellers have perfect predictive knowledge of travel costs/times, does TTV then bring with it a contradiction, since drivers could not possibly anticipate the times they will experience when they are making their route choice decisions? While several authors have argued for such radical departures from accepted methods, it seems that even in such situations it is still possible to extend and utilise our existing tools: this is an issue to which we shall return in subsequent sections.

A much simpler way of interpreting the role of TTV, and the one that has been adopted in previous DfT-funded work, is that essentially TTV is just another attribute to include in the list of components of drivers’ generalised costs. That is to say, there is some measure of TTV (say, the travel time standard deviation on the alternative routes/links) that drivers are assumed to perceive from their repeated use of the transport system, and that the issue then is only how to ‘value’ this additional attribute. This sort of thinking allows us to stay within our conventional equilibrium modelling framework, although (as we discuss below) there are still some significant algorithmic challenges that then arise. Clearly there are also empirical questions as to what extent drivers value and respond to such variability when making their route choice decision. Certainly, though, it may be supposed that particular routes with a high level of TTV (for example, one including a non-priority junction at the intersection with a major road) might be avoided even if their average performance is acceptable, making the case that allowing for TTV should improve our modelling of drivers’ route choice decisions.
4.4 Towards ‘Implementable’ Models

4.4.1 Background

The study of drivers’ responses to travel time variability (TTV) falls within the wider area of network reliability analysis, an international research theme (e.g. the recent third International Symposium on Network Reliability, www.instr2007.com, covering many Asian and European initiatives, but quite distinct initiatives on this topic exist in the U.S.). These efforts have revealed a wide range of theoretical and computational methods that may be implemented in state-of-the-art models/algorithms for traffic assignment and simulation. Despite these advances, such research has yet to feed through into the practice of network assessment. In the past there has typically been a time-lag of several decades for theory to feed into practice (for example, the lag in adopting the basic theory for traffic assignment and for micro-simulation, developed in the 1950s).

However, there exists a more fundamental barrier to the transfer of such knowledge, in that the state-of-practice software tools for network assignment and simulation are much more restrictive than the state-of-the-art methods used in research projects. Many of the new techniques, while combined with a familiar theory (e.g. equilibrium assignment) are not implementable in standard software. While some of the early DfT-funded work on TTV made some progress in adapting particular software tools for such purposes (e.g. Gordon et al, 2001), a much more preferable output will be generic guidance that users can subsequently implement themselves in any standard network assessment software. Also, in representing the actual variation in TTV, it makes sense if the approach can build on the series of previous and on-going DfT-funded studies in this area.

4.4.2 Desirable properties for an implementable model: Fixed trip matrix

The first question to address is: what specifically do we require of the approach when seeking an implementable model in the meaning described above, i.e. what desirable properties might we require and what constraints does this place on our modelling approach? It is helpful to consider these in two stages—firstly the fixed demand case (which is already complex enough), and secondly the variable demand case—since the issues are more readily appreciated in that way, in particular because standard network assignment methods differ in the way they deal with variable demand.

Focusing on the fixed demand case, the only response from drivers (to unreliability or to a network change) that is modelled is route choice: the total demand per OD movement and the departure rates are all assumed fixed. A set of desirable properties are now defined, with subsequent discussion of why they are desirable:

P1 The representation of TTV on the supply-side is in the form of link-based components of travel time variation, e.g. link travel time standard deviations, link travel time probability distributions, correlations between link travel times.
There are a number of points here. Firstly, this suggests a composite approach to modelling actual TTV, as opposed to a cause-based approach (see section 4.2). The reason for selecting such an approach is that, from an empirical viewpoint, it seems a sensible point to start, and it has the chance to build on the series of past and on-going DfT studies in this area that are largely taking the composite variability approach. A second point is that we do not restrict attention only to link standard deviations, but also potentially other moments or even the whole pdf, as well as correlations with other links. This makes the approach more general, and in particular should provide an opportunity at a later stage to incorporate several of the cause-based variability approaches. A third point is that the data on variability is not specified at a route level. There are several advantages to such an approach: it can be argued that surveillance data would yield input data that is most naturally collected at the link level; by inferring route level variability from the link level components, the natural correlations that logically occur between overlapping routes are automatically reflected. Also, the method is robust in that if a network change is made to a link (eg link removal/addition), then the assumptions are still consistent. Note that this does not preclude the link variances/distributions being dependent on link flow, and therefore one can in principle have policy-dependent levels of link variance (assuming the policy affects link flows).

P2 The supply-side TTV representation should be decomposable, in the sense of not being sensitive to arbitrary choices of network definition.

This is a separate point to property P1, and is best understood with an example that satisfies P1 but violates P2. Suppose each link \( a = 1,2,\ldots \) of a network has mean travel time \( \mu_a \) and variance \( \sigma_a^2 \). For convenience, as we may not have access to data on travel time variance across all links of a network, as we described in section 4.2 a typical assumption adopted is that \( \sigma_a = k \mu_a \) where \( k>0 \) is a ‘global’ coefficient of variation assumed to apply to the whole network. Now, consider one particular link, and suppose that it is divided in two in some arbitrary way, but such that the mean travel time is preserved. So if the original link had mean travel time \( \mu \), and the two new links had travel times \( \mu_1 \) and \( \mu_2 \), then we would require \( \mu = \mu_1 + \mu_2 \). By the assumption that \( k \) acts as a global coefficient of variation, the new link should have a travel time variance of \( \sigma^2 = (k\mu)^2 = (k(\mu_1 + \mu_2))^2 = k^2(\mu_1 + \mu_2)^2 \). However, this is inconsistent with the total variance of the two individual links, which is \( \sigma_1^2 + \sigma_2^2 = k^2 \mu_1^2 + k^2 \mu_2^2 = k^2(\mu_1^2 + \mu_2^2) < k^2(\mu_1 + \mu_2)^2 \).

This is an issue of more than theoretical concern, since in practice the choice of which roads are ‘significant’ and therefore to include in a network model is somewhat subjective; inclusion of any additional link often requires a new point of connection to be made to a network (e.g. including a more minor road by adding a priority-junction to an existing major road), and thus often results in splitting of the link with which it intersects.

Alternative solutions to this problem which allow P2 to be satisfied include:

---

\( ^8 \) Typically the mean travel times would be something estimated by the traffic assignment model and therefore be flow-dependent, but this is immaterial to the point being made, and so is not reflected in the notation.

57
– Requiring the model to permit a separate definition of a constant travel time variance for each link, not linked to the mean link travel time, which can be estimated from current conditions and is assumed unrelated to flow or policy interventions.
– Assuming a global variance-to-mean ratio, rather than global coefficient of variation, so that $\sigma_a^2 = k \mu_a$ for each link $a$ (but is there empirical evidence to support this?).
– Specifying the variance of travel time on each link as a separate function of flows, allowing it to be policy sensitive but not directly linked to mean travel time.

This issue of decomposability does not appear to have been one that has motivated many previous Department-funded studies in this area, but seems important for transferability and consistency of approaches.

**P3** The impact of TTV, as perceived by travellers, should be represented as a route-level disutility that is a *linear* combination of route TTV with mean route travel time and other components (route distance, tolls, etc).

While models of decision-making under risk/uncertainty may be typically associated with non-linear functions of expected travel time, traffic assignment models and software is conventionally based on the premise of linear combinations of mean travel time and other attributes. Therefore a practical model would be expected to be of such a linear form. Such a model could take many different forms, however. For example, if we suppose that the only attributes of route generalised disutility/cost $c$ (measured in pence) are route toll $\tau$, route mean travel time $\mu$ and a route TTV attribute, we might adopt $c = \tau + a \mu + \beta \sigma$ or $c = \tau + a \mu + \beta \sigma^2$ or $c = \tau + a \mu + \beta$.lateness, for example, with the premise that constant valuations may be estimated (and the functional forms supported by empirical evidence) for any of the interpretations of $\beta$ adopted.

**P4** The impact of TTV on route level disutility/cost should be *linear-additive* in link-level TTV components.

It should be noted that this property, like property P3, is concerned with the perception of travel cost to the driver, whereas properties P1 and P2 were concerned with the way in which actual variations in travel time are specified to the model. As an example, it is possible to satisfy all of properties P1, P2 and P3, but fail to satisfy P4: travel time variability could be specified by link-level variances in order to satisfy P1 and P2, travel disutility could be assumed to be of the form $c = \tau + a \mu + \beta \sigma$ (satisfying P3), yet this disutility is not linear-additive in link level components.

Much previous work by DfT in this context has indeed been based on the notion of a generalised cost definition comprising a linear combination of travel time mean and travel time standard deviation. As we have seen, this has been justified by a connection to scheduling and late arrival penalties under specific assumptions about the travel time
distribution, but perhaps more convincingly it has been chosen due to its (conceptual) simplicity.

However, in a practical network context, the cost of travel on a route consisting of (say) \( n \) links with link travel time means and standard deviations \( (\mu_i, \sigma_i) \) \( (i = 1,2,\ldots,n) \) would be of the form:

\[
\text{generalised route cost} = a \sum_{i=1}^{n} \mu_i + b \sqrt{\sum_{i=1}^{n} \sigma_i^2}
\]

which is not additive in the *link* cost components \( a\mu_i + b\sigma_i \) \( (i = 1,2,\ldots,n) \), unless one of two things happens:

- \( b = 0 \) (the standard traffic assignment problem without reliability), or
- only one of the \( \sigma_i \neq 0 \) (each route consists of only one variable link).

Note that even the standard assumption of a constant CV \( k \) across the network does not help with this issue: for if we let \( \sigma_i = k\mu_i \) and substitute in the equation above:

\[
\text{generalised route cost} = a \sum_{i=1}^{n} \mu_i + bk \sqrt{\sum_{i=1}^{n} \mu_i^2}
\]

which while now only depending on one link attribute \( \mu_i \), it is still non-additive in that attribute.

Thus this turns out to be a rather serious restriction: nonetheless, it is probably one of the most important properties for implementation in existing network assignment software. This is due to the fact that a key building block of equilibrium traffic assignment algorithms is the ability to perform fast shortest path calculations, even on large networks, using tree-building algorithms. The shortest path algorithms implemented are fundamentally based on the premise that the criterion to be minimised decomposes into a sum of link-level components (as distance and mean travel time naturally do). Moreover, the traditional formulation of static assignment as a convex minimisation problem (used in the classical Frank-Wolfe algorithm) is also based on the premise of linear-additive disutilities across links.

If, instead, we abandon this restriction, then we have what is termed a *non-additive traffic assignment problem* to solve\(^9\), a problem that is still soluble and on which there has been some recent research advances, but which is not yet reflected in commercial network assignment software available to practitioners. To be clear, this is not a major conceptual leap in theory, since the established underlying theory of equilibrium assignment still

\(^9\) It is noted in passing that another topical problem may lead to a non-additive problem, namely in certain forms of road pricing system (e.g. on an urban level the current London scheme, on a wider level distance-based charging methods). That is to say, there seem to be several reasons for more effort in developing and transferring methods for non-additive problems from theory into practice.
holds good: the problem is just that new solution techniques are needed, and such
techniques will typically have to work more with path-based representations of the
network, for which intelligent methods are required to make the methods computationally
feasible for large networks.

An important point to appreciate is that in respect of P4, adopting what is apparently a
simple model of the form \( c = \tau + \alpha \mu + \beta \sigma \) is in fact no more attractive than adopting a
model in which the disutility is linked to some more complex model of scheduling. Thus,
while a linear formulation is generally compatible with demand/valuation approaches,
and since a link-based approach makes sense as it is where data is likely to come from,
the resulting problem does not have as simple a structure as one might expect and hope
for.

**P5** The resulting equilibrium model should satisfy the *convexity* conditions needed to
apply algorithms such as Frank-Wolfe.

This condition is mentioned for completeness, but given that P4 is already satisfied, P5
does not add a great restriction. Assuming P4 to hold, a sufficient condition for P5 would
then be that the link level disutilities are continuous and monotonically increasing in the
total flow on that link.

**P6** Heterogeneity across the driver population in their valuations of TTV (and other
attributes) should be represented in *discrete* user groups.

In the area of demand modelling via discrete choice models, there have been considerable
advances in representing such features via so-called mixed/kernel logit models, where the
heterogeneity may be parameterised by a probability distribution, with valuations
assumed to be randomly distributed across the population. While there has been some
limited, parallel work in the research literature on transferring these methods into
assignment models, such work has not fed through into the assignment software used in
practice. Yet such software does usually include a feature for defining multiple user
classes of traveller, where drivers may be divided into discrete groups each with its own
valuations and OD demand matrix. Through judicious definition of such groups, one may
represent important features of TTV modelling, such as the impact of: different journey
purposes, which are highly likely to have different arrival time constraints and therefore
differing values of TTV; socio-economic group, to reflect (say) the influence of income
on the response to tolls versus TTV; and even to accommodate some departures from
linearity in property P3, such as differential response by journey length (over and above
the attributes modelled – i.e. distance already likely to be included as a linear attribute).

4.4.3 Desirable properties for an implementable model: Variable trip matrix

Here we briefly consider what additional considerations arise when we relax the
restriction of fixed trip matrix.
There are two main groups of approach to combining variable demand and traffic assignment that have been pursued in the context of UK practice: ‘integrated’ models and ‘shell’ models. As the considerations are somewhat different, it makes sense to consider each of these groups of approach in turn.

**Integrated models** aim to represent the problem of finding an equilibrium demand and traffic assignment solution in a single, combined form. The classical example of such an approach is the elastic demand traffic assignment model, with the demand response represented by elasticity between a travel/no-travel option (per OD pair and per user class), which under mild assumptions may be integrated with the traffic assignment equilibrium as a single convex minimisation problem, solvable by standard means such as the Frank-Wolfe algorithm (or, typically, some specialised variant thereof). Such approaches may be extended to other forms of demand response\(^\text{10}\), such as joint distribution and assignment, and some limited forms of joint departure time and assignment to multiple time periods, many of which are available (for example) in the SATURN software.

In such models, the primary TTV considerations are governed by the fixed demand requirements, already discussed in section 2.1. However, since the model will use a single definition of generalised cost/disutility for all choice dimensions, any assumptions made with regard to the route choice dimension in section 2.1 make implicit requirements on the demand-side.

Regarding the remaining properties, while the pure supply-side requirements P1 and P2 are the same in the variable demand case, P6 now extends to a discrete form of heterogeneity in the elasticity parameters used for any demand model. Property P5 is particularly serious in the context of problems such a departure time choice, where queue spillover may cause ‘asymmetric’ interactions in the travel time relationships, and therefore violate the convexity conditions; otherwise, standard forms of elastic demand function may be accommodate in the theory, where the alternative costs are assumed flow-independent. Assumption P4 only becomes more serious in the context of departure time choice and dynamic assignment (e.g. via CONTRAM), but such technicalities are considered to be beyond the scope of the present note.

Turning attention, secondly, to what we have termed **shell approaches**, the idea behind such methods is that demand responses may be handled by a separate model that is external to the traffic assignment software (the term ‘shell’ being motivated by the idea of a ‘software shell’), but which can both feed demand levels into the assignment model and received feedback on generalised costs from the assignment model. A prime example of such an approach is DIADEM. In such approaches it is possible to consider the traffic assignment requirements as distinct and separate from the demand model requirements.

For example, it may be that the demand model represents departure time choice through a model that explicitly considers scheduling considerations, whereas the traffic assignment

\(^{10}\) among others, they are compatible with the hierarchical logit formulation which is widely used for variable demand problems
model represents route choice through a linear composition of say mean and standard deviation. While this is possible in principle, it would require some careful handling of the feedback of generalised costs from the assignment model to the demand model, since the two models may be defining them differently. Such an approach may also cause difficulties at the level of economic appraisal, which is the common end-point of such network studies, though such inconsistencies (between the demand, network and appraisal modelling) are commonly accepted in transport modelling practice.

4.4.4 Implications and conclusions

Based on the requirements presented in section 4.4.3, the question arises as to what classes of model might satisfy these requirements? The answer seems to be very few, but at least one model does exist, as we now define. The properties of such a model are:

- Route disutility assumed to be a linear combination of travel mean and variance.
- Travel time variability specified by link-specific variances that are independent of flow.
- Differences in response to TTV across the population represented by multiple user classes.

This model satisfies all of the requirements specified, and for this reason there seems to be value in at least considering its potential. We can therefore pose some interesting questions. Is there any empirical evidence in support of such a model, if judicious use of market segmentation (user classes) is made? Aside from its consistency with current theory and practice, are there any other theoretical arguments in support of it? How far can the model above be extended?

On this latter point, for example, the second bullet point above can be said to be unrealistic since it has already been established in several empirical studies that journeys at low flow levels are typically highly reliable, whereas as flows increase towards capacity TTV becomes more of an issue. If travel time variances can be assumed to be related to travel time means, then the second requirement could be dropped, since then our problem is effectively a simple extension of existing methods, through an adaptation of the link travel-time flow functions; if not, then it would seem to require new modelling tools that are not so commonly found in practice (see, for example, Mirchandani & Soroush, (1987); Watling, (2002)).

If such a model is not adopted, then it seems that we must drop one or other of the requirements given in this section. For example, in order to support a model that uses a linear combination of mean and standard deviation we must drop property P4, which has enormous implications for implementing the model in current software. On the other hand, if we do decide to drop P4, then the question arises: why would we choose to use the mean/standard-deviation model? That is to say, since it offers nothing in the way of easing immediate implementation, it seems that once property P4 is dropped, the
discussion ought to open up to consider a much wider range of potential modelling approaches, examples of which are discussed in the following section.

4.5 Widening the Range of Models

In this section, by using two example papers/methods, we discuss some of the wider issues that are relevant to the introduction of TTV to traffic assignment models. The first method (Watling (2006)) was initially motivated by the author’s involvement in a DfT study of travel time variability around 2001 (reported in a paper by Gordon et al. (2001)), and earlier work by Noland and colleagues (Noland et al. (1998)). Rather than being a completed work, it is more of a think-piece on how to go further in implementing notions of risk in equilibrium traffic assignment models.

As there seems to be no simple approach that is both justified by some kind of theory and implementable in existing assignment software, it seems sensible to go back first to specifying a model and then seeking to improve it with reliability capability. The Watling (2006) paper effectively combines approaches already proposed to provide such an approach. It combines:

- An explicit link to scheduling through the classical Vickrey-type late arrival penalty model discussed in Chapter 2.
- A link-based model of travel time variability from which route-level variability can naturally be derived. Since the available variability data is often at a link level, this is an attractive property, in contrast to some papers that postulate route level distributions that are not consistent with any link-level definition.
- Correlations between paths due to overlapping links: another consequence of starting with a link component definition of variability is that one then gets natural correlation between routes that share links.

It can be extended to cases such as multiple user classes (say with different levels of risk aversion, perhaps representing different journey purposes and/or income levels), and to skewed travel time distributions (e.g. to represent incident effects). So it has a quite consistent link to much of the behavioural theory about travel decision making under uncertainty, while it should be possible to calibrate it to monitoring data on travel times.

The disadvantage is that a) it needs information on preferred arrival times, and b) it is still non-additive. On the latter point, there are, however, several approaches to dealing with non-additive equilibrium assignment problems, some of which are currently under investigation in a study at PhD at Leeds.

While the model presented above is not the only model in the literature on dealing with traffic assignment under stochastic travel times (and it could be interesting to review and compare others), it has particular relevance to DfT/UK work in this area due to its close connection/derivation from the approaches that have been pursued here. Although it
needs new algorithms, it does not go away from state-of-practice principles such as equilibrium assignment.

The second paper considered (Clark & Watling, 2005) illustrates a somewhat different perspective on the problem of travel time reliability. The work above, and much of DfT work, has focused on the traveller’s perspective of such risk/variability and how this will impact on average economic benefit. But then the planners themselves may also take a view on reliability and aim to introduce measures that mitigate its effects. This is a quite separate issue: even if the individual travellers ignore travel time reliability when making their route choice decisions, it is still the case that real travel times are actually variable, which means that the economic benefits themselves are variable. The Clark & Watling paper, examines the problem of network unreliability from just such a perspective.

The method has the advantage that it can be implemented as an external ‘shell’ to existing traffic assignment software. The aim is to link the variability in actual travel times to one particular source of unreliability, namely that caused by day-to-day variability in the OD demands and traveller decisions, which themselves lead to variability in link traffic flows. Based on assumptions of Poisson distributed OD demands (stochastic but inelastic), a method is deduced that examines the distribution of total travel time in the network. The method differs from methods previously proposed in that it requires no Monte Carlo simulation (so is fast and reproducible), and indeed needs only a single run of a traffic assignment model (with conventionally available data) in order to operate. A particular contribution of this work is to examine the effect on the full probability distribution of total travel time, which we may typically expect to be positively skewed due to the asymmetric effect of ‘bad days’ relative to ‘good days’.

Again, this is not the only approach proposed in the literature to examine the planner’s view of network unreliability, but has particular attraction as it can be implemented with any static assignment package that is able to output route information, is computationally highly efficient, can be combined with any behavioural model describing the travellers’ responses to unreliability, and can be implemented without any new data.

Both approaches described above assume demand to be inelastic, so it is also pertinent to ask to what extent they could be integrated into a feedback system with demand modelling. Many approaches would be feasible. With a “shell” approach to modelling the demand response, and adopting the Watling (2006) late arrival penalty approach, the assignment model needs to feedback OD equilibrium travel costs/disutilities to the demand model. As in a conventional equilibrium assignment model, the late arrival penalty approach would naturally produce such outputs, the only difference being that now the equilibrium costs/disutilities would include the penalty for late arrival. If the demand model would prefer as input OD mean travel times, say, or even OD travel time variances, then these could also be estimated. In these latter cases, since alternative used routes on a given OD pair would have equal disutility but generally different travel times means and variances, one would feedback such values flow-weighted and averaged over all used paths. This is analogous to, say, a conventional traffic assignment based on a combination of travel time and distance, where OD travel times at equilibrium may be
estimated over used paths even though not equalised (since the combination of travel time and distance is equalised). Alternatively the late arrival penalty model could be combined with an integrated demand model.

Considering the Clark & Watling (2005) approach, at first sight this would seem a little more onerous to extend, since it challenges the traditional notion of deterministic networks and single-valued appraisal measures. Here the main difficulty with a shell approach would be the limitations of the external shell, in still being intrinsically connected to deterministic notions of appraisal: it is here that the work would be required. An alternative again would be to consider integrating the demand response into the assignment model, though the feasibility of this would need to be further examined. Specifically on this issue, the source paper primarily proposes an analytic procedure (i.e. one that does not require Monte Carlo simulation) for implementing the approach, but also makes comparisons with a second procedure that exploits Monte Carlo simulation. The latter, Monte Carlo simulation approach has the advantage of offering greater flexibility (and still would require only the run of a single equilibrium traffic assignment model), and it is this approach that would be most readily extendable to the elastic demand case. While some thought is needed about the most appropriate modelling assumptions for demand which is now both elastic and stochastic (variable between days), there is no reason why such an approach could not be implemented, with probability distributions now estimated of multi-modal measures of economic benefit (rather than single-valued measures).

One of the main computational hurdles in implementing such an approach is the potentially time-consuming nature of Monte Carlo simulation, but in parallel approaches we have found there to be dramatic reductions possible in computational effort through the use of intelligent Monte Carlo methods (Watling et al, (2004); Sumalee & Watling, (2007)).

Overall, it is clear that there remain substantial challenges, but there are a number of current efforts which are promising. It would be sensible to construct a work programme which sets out the requirements in a clear way, and investigates the feasibility of alternative approaches, with a view to developing practical approximations. Such a work programme should also compare the results against the methods in the current Guidance.

4.6 Public transport supply issues

As was made clear at the outset, the focus in this section has been on highway network models for modelling private vehicle travel, and it is therefore pertinent to ask whether similar considerations are appropriate in the case of modelling bus or rail networks. It is probably fair to say that the development of modelling tools for representing detailed networks in such cases has lagged behind the rather advanced development of highway network modelling. This is also reflected in practice, yet it is now becoming more common to see assessments that involve public transport networks. For example, assessments especially of bus networks involving aggregate, frequency-based approaches
such as EMME2, or even line-based simulation approaches, are now increasingly seen. In 
the public transport case, however, it is not so easy to identify a common theoretical 
paradigm (analogous to equilibrium assignment in the highway case) to which we can 
refer in general terms. For high-frequency services, perhaps the closest to such a 
paradigm is the ‘optimal strategies’ approach developed by Spiess & Florian (1989) that 
underlies EMME2, one of the most commonly used packages in practice.

The optimal strategies approach is based on the premise that, in such high frequency 
cases and especially when a trip involves interchange, travellers do not make a choice of 
a particular bus at a particular time (or a particular sequence of buses at particular times). 
Rather they make a choice of strategy. A strategy could be, for example, ‘go to the first 
bus stop and take the first out of bus numbers 3 and 5 to arrive there, get off in the city 
centre, and then take the first of bus numbers 8, 9 or 10 to depart’. In this definition, there 
is an intrinsic notion of unreliability (the travellers do not know which bus will arrive first 
due to potential unexpected delays to buses upstream, or because they are not sure of 
their arrival time at their first stop, or because they are unsure of their arrival time at an 
interchange point). Thus it is assumed that travellers mitigate the impacts of unexpected 
variability by adopting such flexible strategies, and thus when they evaluate the total 
expected OD travel time associated with a particular strategy, they are assumed to already 
account for such unexpected delays in their calculation. This is quite different to the 
standard approach in highway assignment models (though there has been research on 
methods for representing en route diversion in response to unexpected delays/queues, 
perhaps the most similar phenomenon). It means that some elements of TTV are already 
accounted for in standard approaches to public transport assignment.

On the other hand, it should be recognised that state-of-practice methods such as those 
above are also somewhat limited, particular in the way in which they deal with vehicle 
capacity constraints (which affects both the mean and variance in waiting times) and 
travel choices that are based on more than just expected OD travel time (see for example 
Teklu et al, (2007)), both of which have relevance in a TTV context: For consistency, one 
would also note that road-based public transport modes, unless furnished with dedicated 
lanes, suffer many of the same consequences of TTV as private cars, the variability in 
private car demands and flows therefore having an impact on private transport TTV. 
While there do exist some approaches for consistent treatment of multiple modes in a 
network context (e.g. Uchida et al, (2007)), their extension to TTV issues is an area in 
need of further research.

4.7 Concluding remarks

The supply model is essential if we are to incorporate TTV in a general modelling 
framework. In the description of approaches above, we have made a classification into 
two broad areas: firstly, how to represent actual TTV, and secondly, how to model the 
effect of TTV on network performance. On the first issue of representing actual TTV, 
there are a range of prominent Department-funded studies that have been completed or 
are on-going. The approaches may be divided into those (the majority) that look at the
composite effect on TTV of all sources, and those that aim to decompose the variability into its component sources, namely day-to-day-variability of demand and incidents. While many of the studies have focused on standard deviation as sufficient for representing TTV, we have raised the question as to whether fuller information on the complete distribution of travel times is required to obtain a true picture of TTV impacts, but that we may also consider methods for generating the first and second moments of the distribution. This area of investigating and modelling actual TTV is clearly a rather active area of DfT-supported work, and further reviews of developments would be appropriate as new results emerge from on-going studies. However, we believe that thought should be given as to whether the objective of representing actual TTV needs to be widened to summary measures beyond the second moment, given the typically asymmetric impact of flow variations and incidents on the right-hand tail of the travel time distribution.

Turning to the second issue, of incorporating travel time reliability in network assignment models, a number of theoretical and practical challenges remain, and on this side of the problem it seems there is much less DfT-supported activity. To some degree, these challenges echo the earlier debate of chapter 2 concerning the relative merits of the mean vs. variance and scheduling approaches. More fundamentally, however, the consideration of TTV raises a wider question, namely what is the mechanism for transferring advances in network assignment theory into practice?

In section 4.4, we have explored how far existing methods can be taken in this way (based on issuing guidance to users of existing software), and the answer seems to be: a little further but not very far. Looking to widening the scope of modelling tools considered for this purpose, the example methods described in section 4.5 are, we believe, practicable and realistic and do not go so far as to criticise the theoretical foundation of current practice (e.g. concepts such as ‘equilibrium’ are not accepted by all researchers when modelling a variable environment), but they do need specialist software tools that can make the best use of current research and knowledge. The incorporation of travel time reliability within practical transport planning tools is an area that is evolving quickly and will continue to do so, and requires the continued support of the Department to keep apace with the research developments.
5. Integration with Appraisal

5.1 Introduction

The state-of-the-art review of theory, valuation evidence, data, and modelling makes it clear that there are important outstanding issues in the appraisal of schemes/policies which bring about improvements in travel time variability. Only limited progress has been made towards harmonizing the theory across modes (see Chapter 2 above). The valuation studies which have been conducted for car, rail and bus travel, and for freight, are based on various theoretical foundations. Whether the variation and inconsistencies in the empirical results (Chapter 3) are due to this, or to differences in the survey approaches used, or to real differences in sensitivity to variability from one context to another, is not transparent at the present time. Against this background it is challenging to make recommendations on procedures for use in multi-modal appraisal.

We are aware that there is a practical imperative. Failing to analyse variability means leaving variability improvements out of appraisal and decision-making, in direct contradiction both to policy and to the majority of the research evidence. For example, the Eddington Transport Study has indicated that “reliability” improvements are important to UK economic policy and enhancements to modelling and appraisal methods are necessary (Eddington, (2006)). Analysis of the 1998 Roads Review, which used a simple proxy indicator for “reliability” based on flow/capacity, found that Ministers appeared to place great weight on reliability improvements, with an implicit value of £17-29 million per point on the reliability scale (Neutral; Slight; Moderate; or Large improvement - each of these were defined quantitatively) (DETR, (1998); Nellthorp and Mackie, (2000)). Attitudinal research frequently finds that reliability is one of the most serious concerns of transport users (e.g. AVV, (2003); SRA, (2003)).

In addition, the valuation evidence cited in this report leaves little room for doubt that transport users are willing-to-pay something for improvements in reliability – the questions are really 'how much' and 'for which measure of reliability', which takes the discussion back to the empirical and theoretical issues.

Our recommendation is therefore to proceed with the necessary theoretical work, followed by investment in practical network models for the full set of modes. Given the likely expense of developing a new generation of models, we think that this is a better course of action than developing an 'intermediate' generation of models based on the reliability ratio approach as an interim measure. Whilst an intermediate set of models would take some steps forward towards estimating the benefits arising due to improvements in travel time variability, this would also bring some disadvantages. The estimates produced by such models could be substantially erroneous. In addition we believe that if that interim step was taken, it would then become much more difficult to move on to a satisfactory generation of models in, say, 4-6 years' time - since many in the industry would have sunk considerable investments into the 'intermediate' approach and would likely be looking for a return on their investment.
At this point, we note the elements of the appraisal that will need to be renovated (below), and we have provided a rough illustration to DfT of the implications of choosing the reliability ratio approach, for all these steps.

5.2 Implications for incorporating “Reliability” into the Appraisal Procedure

Although the details are not yet agreed, on practical grounds the first stage is to devise an appropriate metric for incorporating a measure of Travel Time Variability into the definition of Generalised Cost. Some variations by mode could be considered.

The second stage is to agree on a range of segment-specific monetary valuations of the chosen metric. Again, some variations by mode could be considered.

Having identified how the generalised cost function can be extended and how reliability improvements can be valued, it will then be necessary to give consideration to downstream aspects of reliability appraisal, including:

– user benefit estimation;
– operator revenue;
– the public accounts;
– wider economic benefits;
– growth of reliability values over time;
– deriving a Present Value of Benefits for reliability;
– reporting and presentation of results.
6 Conclusions

Two versions of the theoretical paradigm for representing travel time variability have been presented here, the mean vs. variance approach, and the scheduling approach. In both cases, a strong recommendation is that further work is needed to progress the theoretical basis and to test the hypothesized model forms.

In our view, it is not currently possible to choose between these two versions without a clearer understanding of their similarities and differences. Whilst the mean vs. variance approach, together with the associated metric of the ‘reliability ratio’, has the advantage of greater simplicity, our belief is that the scheduling approach may reveal important, and unique, insights into behavioural responses to unreliability, particularly in the context of scheduled (but infrequent) public transport services.

If there is to be a meaningful comparison of travel time variability across modes, then any remaining challenges to the correspondence between the mean vs. variance and scheduling theoretical approaches must be resolved.

The review of empirical evidence on values for variability has highlighted the existence of a growing body of work on the modelling of the valuation of travel time variability. Not all existing work is available in English and a tranche of work may not even be publicly available at all.

While all reviewed studies agree that variability is a factor of substantial importance, there are no generally accepted monetary values for variability, or indeed a reliable estimate of the relative weight of travel time and travel time variability. It should be acknowledged that the valuations of variability in the present literature come from very specific investigations, and are not used in cost-benefit analyses even in the respective countries of origin. Often, the studies are relatively small scale, and in some cases, variability is not the main topic of investigation.

A recommendation from this research is that there is a need for a major new study into the valuations of variability, investigating both the question of how information on variability is presented to respondents and how it is later specified in econometric models that are estimated on the data. This must, however, follow on from agreement on the theoretical framework for journey time variability. The aim should be to achieve a high level of consistency between the underlying theory, the data collected and the estimated model.

The supply model is essential if travel time variability is to be incorporated in a general modelling framework. A classification is made here into two broad areas: firstly, how to represent actual TTV, and secondly, how to model the effect of TTV on network performance. On the first of these, the approaches may be divided into those (the majority) that look at the composite effect on TTV of all sources, and those that aim to decompose the variability into its component sources, namely day-to-day-variability of...
demand and incidents. Whilst much previous research has assumed that the standard deviation is sufficient for representing TTV, we believe that thought should be given as to whether the objective of representing actual TTV needs to be widened to summary measures beyond the second moment, given the typically asymmetric impact of flow variations and incidents on the right-hand tail of the travel time distribution.

With respect to incorporating travel time reliability in network assignment models, a number of theoretical and practical challenges remain, mainly echoing the issues concerning the relative merits of the mean vs. variance and scheduling approaches. A more fundamental issue remains, however: namely, the mechanism for transferring advances in network assignment theory into practice. For further advances in the practice of TTV modelling to take place, the practice of issuing guidance for use in a small number of accepted commercial packages may need to be reconsidered - the use of specialist software tools that can make the best use of current research and knowledge should also be supported by the Department.

The state-of-the-art review of theory, valuation evidence, data, and modelling makes it clear that there are important outstanding issues in the appraisal of schemes/policies which bring about improvements in travel time variability. Without resolution of these issues, at this point we do not consider it appropriate to do more than list the elements of the appraisal that will need to be revised. Any further elaboration is embryonic without resolution of the underlying theory.
Annex I: Development of theory (Bates, Polak, Jones and Cook, 2001)

The following is largely based on the article by Bates, Polak, Jones & Cook in Transportation Research E (April/July 2001). It draws on work by Noland and Small (1995) [N&S], which in turn builds on the earlier theoretical contributions of Gaver (1968), Polak (1987) and Small (1982).

The central notion is that travellers receive a disutility not only from travel time but also from arriving at their destination either earlier or later than desired. This is essentially the concept of schedule delay formulated by Small (1982), based on earlier work by Gaver (1968) and Vickrey (1969).

The standard assumption relating to discrete choice is that the traveller will attach a utility to each of the options being assessed, and choose that with the highest value (or, equivalently, that with the lowest generalised cost).

We take PAT as the preferred arrival time and the issue is then to choose the time of departure from home, t_h. We assume that for any possible t_h there is an associated utility U(t_h;PAT) which depends on the preferred arrival time. The rule for departure time choice is then

\[ \text{Max } U(t_h; \text{PAT}) \text{ with respect to } t_h. \]

Without being explicit about the form of the utility function, it is reasonable to expect certain quantities to play a role:

- the total travel time T (this may well depend on when the journey begins, so that \( T = T(t_h) \));
- whether the arrival time is before or after (or equal to) the preferred arrival time: this depends on the sign of the quantity \( \text{PAT} - (t_h + T(t_h)) \);
- the amount by which the actual arrival time differs from the preferred arrival time: this is given by the absolute value of the quantity \( \text{PAT} - (t_h + T(t_h)) \).

There may also be money consequences of travelling at different times (eg time-dependent road user charges, peak-period pricing on public transport), though for simplicity we will ignore these.

Empirical research (e.g., Small, 1982) suggests a possible travel (dis-)utility for the trip (NB excluding utility associated with the destination) of:

\[ U(t_h) = \alpha T + \beta \text{SDE} + \gamma \text{SDL} + \theta D_L \]  

(I.1)

where,
SDE is early schedule delay defined as $\text{Max}(0, \text{PAT} - (t_h + T(t_h)))$
SDL is late schedule delay defined as $\text{Max}(0,(t_h + T(t_h)) - \text{PAT})$
$D_t$ is a dummy variable which is equal to 1 if SDL > 0 and 0 otherwise
and $\alpha, \beta, \gamma$ and $\theta$ are model parameters, assumed to be negative.

We refer to the contributions to utility other than the pure disutility due to travel time $T$ as "schedule disutility".

Although this utility specification is plausible, a number of other variants can be conceived, eg a "band of indifference" around the preferred arrival time so that, for example, no schedule disutility is incurred provided the arrival is within five minutes of the preferred arrival time $\text{PAT}$ (eg Mahmassani & Chang (1986)).

If there are no capacity restrictions, then we do not expect $T$ to vary with $t_h$: consequently, under conditions of certainty, all travellers will choose $t_h$ so as to arrive exactly at $\text{PAT}$. However, if capacity is limited, then this is no longer possible. To make the capacity effects explicit, N&S (1995) initially split the travel time $T$ into the sum of two components ($T_f + T_x(t_h)$), where

- $T_f$ is the (fixed) free flow travel time between home and work
- $T_x$ is the extra travel time due to recurrent congestion, which is assumed to be a deterministic function of the departure time

We now introduce travel time variability by defining an additional element of travel time $T_r$ which is a random quantity greater than or equal to zero. The distribution of $T_r$ is in general dependent on $t_h$, so we can write it as $f(T_r|t_h)$. Thus, total travel time $T$ is now composed of:

$$T = T_f + T_x(t_h) + T_r(t_h). \quad (I.2)$$

Hence, for a given departure time $t_h$, the traveller will arrive late if:

$$t_h + T_f + T_x(t_h) + T_r(t_h) > \text{PAT}$$

Since the first three terms are (assumed) non-stochastic, and $T_r$ is assumed not to be negative, it can be assumed that travellers will not set out later than $\text{PAT} - (T_f + T_x(t_h))$, ie

$$t_h \leq \text{PAT} - (T_f + T_x(t_h))$$

since otherwise they would certainly be late. For convenience, we will write:

$$H(t_h) = \text{PAT} - (t_h + T_f + T_x(t_h)) \geq 0 \quad (I.3)$$

Once we allow for $T_r$, it becomes possible to speak of the probability of late arrival, which we write as $p_l(t_h)$. The traveller is now faced with a non-zero probability of late arrival, where the probability is based on his (subjective) estimate of the probability
distribution function of $T_r(t_h)$. The optimum value of the term $H(t_h)$ can be thought of as the extra amount of time (termed by Gaver (1968) the ‘headstart’) which travellers allocate to the journey in virtue of the possible incidence of unpredictable delays – ie, a safety margin (see also Knight, 1973).

The standard theoretical approach to choice problems under uncertainty is that of "maximising expected utility": that is, choose the course of action which, bearing in mind the probabilities of different outcomes, has the highest value of expected utility. This is a development of the classic approach by von Neumann & Morgenstern (1947) - for a useful discussion of choice under uncertainty, see Deaton & Muellbauer (1980: §14).

The approach implies that the traveller needs to assess all the eventualities resulting from different possible outcomes, though in practice, of course, he is likely to adopt a simpler version of this strategy.

Most practical applications of this approach tend to make a large number of restrictive assumptions. For example, they may assume that the utility function is linear, that it is the same for all travellers, and that the distribution of travel time is known. However, the theory is compatible with an individual's definition of utility and an individual's definition of the travel time distribution.

To solve the departure time choice problem on the basis of the Maximum Expected Utility approach, we need to find $t_h^*$ – the value of $t_h$ which maximises $E[U(t_h)]$, where

$$E[U(t_h)] = \int_0^\infty U(t_h) f(T_T | t_h) \, dT_T$$

(1.4)

where the expectation operator $E[.]$ is based on the traveller’s subjective assessment of the distributions of the elements that affect utility.

Because of the random element $T_T$, all the arguments in the utility function are random.

In order to make progress, it is convenient to assume Small’s widely-used utility function, which allows us to write:

$$E[U(t_h)] = \alpha E[T(t_h)] + \beta E[SDE(t_h)] + \gamma E[SDL(t_h)] + \theta p_L(t_h)$$

(1.5)

Although the term $E[T(t_h)] = T_T + T_T(t_h) + E[T_T(t_h)]$ represents straightforwardly the mean travel time, the expected values of SDE and SDL are more complex, because of the discontinuities of the underlying distributions. For a given $t_h$, the probability of late arrival $p_L(t_h)$ is:

$$p_L(t_h) = \int_{H(t_h)}^\infty f(T_T | t_h) \, dT_T$$
and if the individual arrives late, then SDE will be zero. Hence, the distribution of early Schedule Delay \( g(SDE|t_h) \) is:

\[
g(SDE | t_h) = \begin{cases} 
  f((H(t_h) - SDE) | t_h) & \text{if } SDE > 0 \ [so\ that\ T_e < H(t_h)] \\
  p_L(t_h) & \text{if } SDE = 0 \ [so\ that\ T_e \geq H(t_h)]
\end{cases}
\]

The consequence is that \( E[SDE(t_h)] \) is given as

\[
E[SDE(t_h)] = \int_{0}^{H(t_h)} SDE.g(SDE | t_h) d(SDE) + p_L(t_h).0 = \int_{0}^{H(t_h)} SDE.f((Z(t_h) - SDE) | t_h) d(SDE)
\]

Changing the variable of integration from SDE to T, where \( SDE = H(t_h) - T_r \), we have

\[
E[SDE(t_h)] = - \int_{0}^{H(t_h)} [H(t_h) - T_r].f(T_r | t_h) dT_r \\
= (1 - p_L(t_h)). H(t_h) - \int_{0}^{H(t_h)} T_r.f(T_r | t_h) dT_r \quad (I.6a)
\]

The corresponding formulae can be derived for \( E[SDE(t_h)] \):

\[
E[SDL(t_h)] = \int_{0}^{H(t_h)} T_r.f(T_r | t_h) dT_r - p_L(t_h). H(t_h) \quad (I.6b)
\]

Thus, to find the optimum solution \( t^*_h \), we should differentiate Eq I.5 with respect to \( t_h \) and set to zero, which gives the following outcome:

\[
\frac{\partial}{\partial t_h} \left[ \alpha \left( T_f + T_x(t_h) + \int_{0}^{H(t_h)} T_r.f(T_r | t_h) dT_r \right) \\
+ \beta \left( H(t_h).\int_{0}^{H(t_h)} f(T_r | t_h) dT_r - \int_{0}^{H(t_h)} T_r.f(T_r | t_h) dT_r \right) \\
+ \gamma \left( \int_{0}^{H(t_h)} T_r.f(T_r | t_h) dT_r - H(t_h).\int_{0}^{H(t_h)} f(T_r | t_h) dT_r \right) \\
+ \theta \int_{0}^{H(t_h)} f(T_r | t_h) dT_r \right] = 0 \quad (I.7)
\]

To evaluate this, we need to make use for the formula for differential of an integral, which can be written in general terms as:

\[\text{Note that under all circumstances,}\]

\[E[SDL(t_h)] - E[SDE(t_h)] = E[T(t_h)] - (PAT - t_h)\]

\[75\]
\[
\frac{d h(x)}{dx} \int_{a(x)}^{b(x)} f(x,t) dt = b'(x) f(x,b(x)) - a'(x) f(x,a(x)) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t) dt
\]

Note that from Eq (1.3), we have: \( H'(t_h) = -(1 + T'_1(t_h)) \). If \( T_x \) does not vary with \( t_h \), then \( H' = -1 \).

Applying the “differential of an integral” formula, we obtain
\[
\alpha \left( T'_1(t_h) + \int_0^{\infty} T_r f'(T_r | t_h) dT_r \right) \\
+ \beta \left( H'(t_h) \int_0^{H(t_h)} f(T_r | t_h) dT_r + H(t_h) H'(t_h) f(H(t_h) | t_h) + H(t_h) \int_0^{H(t_h)} f'(T_r | t_h) dT_r \\
- H(t_h) H'(t_h) f(H(t_h) | t_h) - \int_0^{H(t_h)} T_r f'(T_r | t_h) dT_r \right) \\
+ \gamma \left( -H'(t_h) H(t_h) f(H(t_h) | t_h) + \int_{H(t_h)}^{\infty} T_r f'(T_r | t_h) dT_r \right) \\
+ \theta \left( -H'(t_h) f(H(t_h) | t_h) + \int_{H(t_h)}^{\infty} f'(T_r | t_h) dT_r \right) = 0
\]

(1.8)

Simplifying and collecting terms:
\[
\alpha T'_1(t_h) + (\alpha - \beta) \int_0^{H(t_h)} T_r f'(T_r | t_h) dT_r + (\alpha + \gamma) \int_{H(t_h)}^{\infty} T_r f'(T_r | t_h) dT_r \\
+ H(t_h) \int_0^{H(t_h)} H'(t_h) f'(T_r | t_h) dT_r - \gamma \int_{H(t_h)}^{\infty} f'(T_r | t_h) dT_r \\
+ H'(t_h) \left[ \beta \left( 1 - p_L(t_h) \right) - \gamma p_L(t_h) \right] \\
+ \theta \left[ \int_{H(t_h)}^{\infty} f'(T_r | t_h) dT_r - H'(t_h) f(H(t_h) | t_h) \right] = 0
\]

(1.9)

In practice the solution \( t^*_h \) which maximises expected utility will depend on:

- the values of the coefficients in the utility function, which express the basic reaction to early or late arrival, together with the disutility of longer travel time;
- the way in which the anticipated extra travel time due to congestion \( T_x \) varies with the departure time;
- the (perceived) variance of the random travel time component \( T_r \), and the way in which it changes with the departure time.

In general, it is not possible to write down an analytical solution to this maximisation problem in terms of Eq (1.9). Most authors have dealt with the simpler case in which the
distribution of the random component is fixed \( \frac{\partial}{\partial t_h} f_h(T_r) = f_h'(T_r) = 0 \) and the component \( T_x \) is independent of the departure time \( (T_x'(t_h) = 0 \) (in which case \( H' = -1 \)). This means that the condition resolves to:

\[
-\left[ \beta[1 - P_L(t_h)] - \gamma P_L(t_h) \right] + \theta \left[ f(H(t_h) | t_h) \right] = 0 \quad \text{whence}
\]

\[
P_L(t_h) \left[ \beta + \gamma \right] + \theta \left[ f(H(t_h) | t_h) \right] = \beta
\]

If, additionally, \( \theta \) is assumed to be zero, implying that there is no disutility in being late \textit{per se}, it can be seen that the optimum probability of arriving late is given by \( p*_{L} = \beta/(\beta + \gamma) \), so that people choose a departure time to balance the consequences of early and late arrival.

In the somewhat more general case discussed by Noland and Small (1995), the distribution of non-recurrent congestion delays \( f(T_r) \) is, as before, assumed not to vary with \( t_h \), but \( T_x \) is. In this case, therefore, the condition is slightly more complicated:

\[
\alpha T_x'(t_h) + H'(t_h) \left[ \beta[1 - P_L(t_h)] - \gamma P_L(t_h) - \theta \cdot f(H(t_h) | t_h) \right] = 0
\]

In the case where \( f(T_r) \) is an exponential density with parameter \( b \), N&S are able to show that the optimal headstart \( H(t_h) \) (from which the optimum departure time can be readily calculated) is given by

\[
H^*(t_h) = b \ln \left[ \frac{\theta + b(\beta + \gamma)}{b(\beta - \alpha \Delta)} \right]
\]  \hspace{1cm} (I.10)

where \( \Delta = -\frac{T_x'}{(1 + T_x')} \) is the rate of change of the profile of recurrent congestion at the traveller’s planned arrival time at work (PAT). Note that to ensure the desirable “first in, first out” (FIFO) condition, we must have \( \Delta < 1 \) (see Noland (1997)).

In equation (I.10), the extra amount of time that needs to be allowed because the time is unpredictable (the headstart \( H \)) depends both upon the magnitude of the unpredictable delays (i.e. the value of \( b \)) and on the magnitude of the scheduling penalties \( \beta, \gamma \) and \( \theta \) as well as on the time of day profile of congestion (\( \Delta \)), if relevant. If travel times do \textbf{not} vary systematically during the period under consideration, then \( \Delta = 0 \), in which case the headstart is always positive i.e., \( t_h^* = \) always less (earlier) than PAT – \((T_r + T_x)\).

At the optimal departure time \( t_h^* \), it may be shown (see N&S) that with the exponential distribution for \( T_r \), the probability of arriving late, \( p*_{L} \), is given by

\[
p*_{L} = \frac{b(\beta - \alpha \Delta)}{\theta + b(\beta + \gamma)}
\]  \hspace{1cm} (I.11)

and that the expected utility of travel, at the optimal departure time is
\[ \text{EU}^* = \alpha(T_t + T_x + b) + \theta p_L^* + b \left\{ \beta \ln \left( \frac{\theta + b(\gamma + \beta)}{b(\beta - \alpha \Delta)} \right) - \frac{\theta (\beta - \alpha \Delta)}{\theta + b(\gamma + \beta)} - \alpha \Delta \right\} \]  \hspace{1cm} (I.12)

In equation (I.12) the first term corresponds with the (negative) contribution to “generalised cost” of the expected travel time, the second term is the contribution of the optimum probability of being late and the third term can be interpreted as the contribution of the expected schedule delay.

A number of authors have shown empirically that the sum of the terms \( \beta E[SDE(t^*_h)] + \gamma E[SDL(t^*_h)] \) is well approximated by \( \Phi(\beta, \gamma) \). \( \sigma \) for a wide range of distributions, where \( \sigma \) is the standard deviation of \( T_r \), and \( \Phi \) can be considered constant for any given combination of \( \beta \) and \( \gamma \). This provides some justification for the widespread use of \( \sigma \) as the relevant component in the utility function to indicate the effect of travel time variability (eg Jackson & Jucker (1981), Black and Towriss (1989), Senna (1994)). Strictly speaking, however, this relies on departure time being continuously variable (as with the car mode). In addition, the value of \( \Phi \) is not independent of the form of distribution assumed.
Annex II: Valuation of travel time reliability

Introduction

This Annex provides supplementary discussion to that provided in Chapter 3 on the evidence on passenger reliability values. The review is organised according to geographic area and this provides a useful basis to interpret the evidence that is provided. (Note that the following discussion elaborates on material presented in summary form in Tables 3.1 and 3.2 of chapter 3 in the main body of the report).

United Kingdom

On the basis of extensive SP research, models for the value of time on UK roads have been estimated (Accent and HCG, 1995), with some of these models including variables for reliability. As an example, results show that doubling the observed chance of an unexpected delay is as bad as 13 minutes extra travel time in the case of commuters, or 20 minutes extra travel time in the case of business travellers. Halving this chance is equivalent to a three minute reduction in travel time for commuters, or five minutes for business travellers. This thus highlights important asymmetries in the value of reliability.

Copley et al. (2002) carried out a SP survey among single occupancy car commuters in Manchester. Among the attributes presented for each alternative were a set of travel times (in a bar chart) and average travel time. These data were used to estimate discrete choice models, where the bar chart information was used to obtain a standard deviation for travel time. The results imply that a minute standard deviation of travel time is valued 1.3 times as much as a minute of travel time itself, with the authors referring to this as the ‘reliability ratio’.

A significant body of work by John Polak and colleagues has recently looked at the issue of model development on the basis of data where respondents are faced with a set of possible travel times (or delays). Other than using information on the mean travel time and standard deviation calculated from such data, modelling analyses often rely on an expected utility theory (EUT) approach within a random utility framework. This approach is for example used in the work of Bates et al. (2001), where respondents are faced with alternatives that each have ten possible early or late delays.

In an EUT framework, this uncertainty is incorporated into the utility specification by making use of a weighted summation across the various possible outcomes. The specification of the utility function for each of these outcomes has an impact on the degree of risk averseness, while the definition of the probabilities associated with the different outcomes allows for departures from EUT. These two issues are discussed by Liu & Polak (2007) and Michea & Polak (2006) respectively, and discussed in more detail in Section 3.2.

Hollander (2005) carried out an SP survey among 244 bus users in York. The respondents were shown two alternatives at a time, each described in terms of fare and ten different...
bars (higher means the journey takes longer), each with a departure and an arrival time. On these data, he estimated both models with a mean and a variance of travel time in the utility specification and scheduling models with mean lateness and another term being the sum of mean travel time and mean earliness. The mean-variance specification was less successful (variance term not significant).

An ongoing study by ITS Leeds, funded by the UK Department for Transport, is investigating the impact of punctuality (i.e. whether a train runs to time) and reliability (i.e. whether a train runs at all) on passenger rail demand. Acknowledging the biases that are sometimes inherent in SP data, the study aspires to place greater emphasis on actual (i.e. revealed and reported) behaviour. The analysis of these data is proceeding along two parallel streams, with dynamic econometric models employed to yield evidence on market elasticities, and discrete choice models employed to yield evidence on monetary valuations.

The econometric analysis is supported by three principal data sources, specifically ticket sales (from the LENNON database), GJT (from MOIRA), and punctuality and reliability (from BUGLE). Guided by the expertise of industry representatives, 248 origin-destination flows were selected that are considered potentially insightful. That is to say, these flows had been subject to non-trivial changes in punctuality and/or reliability (at least in terms of the publicly-reported PPM figures) during the period covered by LENNON, i.e. 2002 onwards. The team assembled the aforementioned data in relation to this sample of flows, on a 13-period basis. The econometric model will be specified as fully as possible, given these data. The number of journeys per flow will be related to the fare (revenue/journey), GJT, and reliability/punctuality, as well as income at the origin (and employment at the destination for season tickets), car ownership and petrol prices. Dynamic model specifications will be tested in order to account for lags in the response to changes in reliability, as well as to changes in other explanatory variables.

The discrete choice analysis is supported by a national passenger survey. This survey combined three elements, namely: retrospective questioning, RP data, and SP data. In a similar manner to before, the team sought assistance from industry representatives in identifying survey locations either where passengers had experience of changing levels of reliability over time, or where passengers had a choice between services with different levels of reliability. In the former situation, travellers were asked whether they had changed their behaviour as a result of changing levels of reliability (i.e. a retrospective question). In the latter situation, travellers’ preferences were elicited between these different services, using both RP and SP. In respect of the SP, a development of Hollander’s (2006) presentation was employed, which itself might be seen as a simplification of the so-called ‘clock face’ presentation adopted by Bates et al. (2001). That is to say, a choice was offered between two services A and B, each of which was described in terms of fare, timetabled journey time, and a journey time distribution based on 5 arbitrary journeys. In an extension to Hollander’s presentation, these 5 trips were expressed not solely in terms of journey time, but also their departure and arrival times relative to the timetable. The data from the passenger survey will be applied to the estimation of discrete choice models.
All data has been collected, and modelling is now underway. The project outputs will be reported during the Spring of 2007.

**Switzerland**

In Switzerland, SP surveys designed by Kay Axhausen and colleagues routinely include a variable giving the percentage of trips that are more than 10 minutes late. This measure is not coupled to journey length, and as such it can be seen to play a bigger role in the case of shorter journeys. A recent example of a study including this attribute in the survey design is the mobility pricing study which looked at the acceptance of toll roads (Vrtic et al. 2006). The results from this study find the sensitivity to this reliability variable to be twice as large for car as for public transport. Furthermore, the results suggest that for car, a one percent decrease in the proportion of trips being delayed by 10 minutes is valued in the same way as a one minute decrease in travel time, while for public transport, a decrease in travel time by one minute is valued more than 50% higher than a decrease in the proportion of late trips by one percent. In turn, a reduction in the proportion of late trips by one percent is valued at 0.5CHF for car travel, and 0.2CHF for public transport.

Switzerland was also the setting for one of the most comprehensive recent studies into the modelling of travel time variability, centred around the PhD work of Arnd König. A detailed description of this work is given in König (2004) and in König & Axhausen (2003). All this material is however in German, and only a relatively brief summary is available in English (König & Axhausen, 2002). The work makes use of a host of different SP surveys, looking at mode choice, route choice and departure time choice. In the mode choice experiment, respondents are given a choice between three alternatives. This contains one train and one car alternative, each with perfect reliability and the same travel time, and one car alternative with a much reduced travel time but reduced travel time reliability, given in terms of the rate of journeys being punctual. In the first two of the route choice surveys (both car and PT versions), respondents are given the choice between a reliable option that costs more, and an option that encounters unreliability a certain number of days per week, where respondents are additionally provided with the average delay time. The final route choice experiment provides respondents with a graphical example, with a choice between two routes, one of which is faster, but which, for a certain number of days per week, encounters a delay by a certain time.

The results from this very comprehensive study are difficult to summarise. However, a few points can be highlighted. Independently of the survey used, the analysis always confirms the importance of reliability in respondents’ choice behaviour. Crucially, the analyses also show that the probability of encountering a delay carries a bigger penalty than the actual duration of the delay. The fact of not arriving on time thus seems more important to travellers than the actual duration of any delay. This observation suggests that the possibility of a delay and the duration of it should be treated separately in modelling analyses. Additionally, the work suggests that it is preferable to specify models on the basis of the probability and length of a delay rather than working with a mean

81
travel time in conjunction with a standard deviation. This finding is however potentially influenced by the specific format used for the SP studies.

The results from the analysis can be summarised in a valuation of CHF34 for an average delay of 60 minutes, where the value is slightly lower for PT than for car. The VTTS measures produced on the same data are roughly 20% lower than the valuations of travel time reliability.

**Netherlands**

Rietveld et al. (2001) found in an SP survey among public transport users in the Netherlands that a 50% reduction in the probability of a delay of two minutes is valued at 32 Eurocents, while the value of time is estimated at 13 Eurocents per minute. Therefore one ‘uncertainty minute’ (travelling one minute longer than expected) is equivalent to 2.4 minutes travel time. This implies that the public transport travellers are risk-averse; if they had been risk-neutral, an uncertainty minute would be equivalent to a minute travel time.

In the Netherlands a major empirical study is underway to measure the value to society of travel time benefits and travel time reliability benefits. The study is commissioned by the Dutch Ministry of Transport to a consortium consisting of Significance, the VU University Amsterdam and John Bates. The resulting values of time (VOTs) and values of reliability (VORs) from the study will be used in cost benefit analyses.

The VORs will be estimated using SP experiments. In the past few years, many researchers have designed formats to present unreliability to respondents in SP experiments. These formats use concepts from statistics like “average travel time”, “travel time variance”, “probability of arriving late”, “histograms”, etc. The big question was whether these fairly advanced concepts are understandable for laymen travellers without a degree in statistics or even higher education. To the best of our knowledge, this issue has never been empirically put to test. It is, however, crucial to gain insight in the quality of the empirical results for VOR’s collected from interviews.

The team has, therefore, designed an in-depth interview among a focus group of 30 respondents (Tseng, Verhoef and van der Hoorn, 2007). The objectives of these face-to-face interviews are as follows:

- Test the respondents’ understanding of different reliability presentation formats
- Investigate the respondents’ assessments of these presentation formats with respect to clarity, ease of handling, and visual attractiveness
- Collect the respondents’ preferences of the presentation formats

In the analyses eight different formats were tested. These are all formats that present five different travel times (mass points) within a single option. Some use bar charts (as In Hollander, 2005), some a clockface presentation (as in Bates et al, 2001) and some only use words. Respondents were stratified according to their education level. There were three groups of questions. First, questions about how respondents conceptualise
unreliability themselves. Do they think in terms of average, minimum, maximum travel time or probability of a delay? And how complicated do they find these concepts? Secondly, respondents were prompted with questions to test whether they gave the “right” answer. Thirdly, there were questions about respondents’ assessments of the eight presentation formats regarding clarity, ease of handling, and visual attractiveness, and which ones were preferred.

The interviews supplied a clear “winner” among the eight formats. This format is not only preferred by a majority of respondents, but also equally by lower and higher educated people. And it is the format where the answers to the test questions indicate that reliability is understood by the respondents in the same way as by the researchers. Perhaps surprisingly this is the format that does not use any graphics, but only describes five equi-probable travel times and their corresponding arrival times in words and numbers. This format will be used in the forthcoming main study.

France

In an SP survey among bus users in France, MVA (2000) found that the importance of one minute of standard deviation of travel time was only a quarter of the value of a minute of in vehicle travel time.

Sweden

Eliasson (2004) reports an SP survey among 600 car drivers in Sweden. One of the experiments asked for the choice between alternatives described in terms of travel costs and travel time range (e.g. 28-52 minutes). Another experiment included the frequency of a large unexpected delay (of a specified size). On this material he obtained a reliability ratio of 0.95 for commuting, 0.3 for business travel and 0.59 for other purposes.

In an expert workshop on the value of reliability (see Hamer et al, 2005) some other Swedish studies were mentioned (Stockholm public transport study, 2001; Swedish railway study, 2004). Key outcomes of these are included in chapter 3 of the main report, Tables 3.1 and 3.2

Australia

The work by David Hensher and colleagues in Sydney typically includes a travel time variability attribute in the design of SP surveys (usually route choice). This attribute gives the amount of time by which respondents can expect the travel time to vary across a number of choice situations, where this variation can be to either side. This thus means that while other attributes are trip specific, the variability attribute refers to a multiple trip scenario. More recent work by Hensher’s group makes use of the probability of on-time arrival instead; no results are as yet available from these studies.

Only a limited amount of published results are available on the valuations of the above mentioned travel time variability attribute, with a summary of results across four studies.
given in Hensher (2007). This reports a willingness to pay for a reduction in travel time variability of AUD4.84 per hour for commuters, and AUD5.02 per hour for non-commuters. Three separate travel time components are used in these surveys, and the weighed VTTS across components is given as AUD18.23 per hour for commuters and AUD14.53 per hour for non-commuters. As such, the willingness to pay for reductions in travel time is considerably larger than the willingness to pay for reductions in travel time variability. Here, it should be mentioned that one possible source for this observation is poor understanding on behalf of respondents of the travel time variability attribute. Indeed, in a number of studies making use of data containing this attribute (e.g. Hess et al, 2006), the attribute obtains very low levels of statistical significance depending on the specification used.

It should also be said that the use of separate travel time components in SP surveys has some relation to the topic of travel time variability. Indeed, it can be argued that respondents’ higher valuations of attributes such as slowed down time and crawl time potentially reflect an inherent belief that these components of travel time also carry with them a higher risk of further delay.

From the work of Hess et al. (2006) on the Sydney data, we can obtain ratios between the sensitivities to travel time variability and free flow time of 0.13 for commuters and 0.11 for non-commuters. When using slowed-down time in the denominator, the ratios decrease further, to 0.1 and 0.09 respectively. These results are thus consistent with those reported by Hensher (2007) on data making use of three travel time components, showing lower sensitivities to the travel time variability attribute, although the ratio is even lower with this dataset.

**United States**

Brownstone and Small (2002) provide a review of a number of studies on road pricing in California. One of these is the work of Lam and Small (2001), who use RP data (travel time measurements) from studies into the route choice for the State Route 91 (SR 91). Here, car drivers can choose between toll lanes and free, but often congested lanes. On the basis of the RP data alone, the values for unreliability, measured as the 90th percentile minus the 50th percentile is 11-14 Euro/hour for males and 28-30 Euro/hour for females. This definition of reliability does not differentiate between expected and unexpected congestion; both forms of variation can influence the degree of unreliability, but the average amount of congestion will be reflected in the median.

Small, Winston and Yan (2002) also use RP data from the SR 91, but in conjunction with SP data. The SP/RP value of unreliability is 26 Euro/hour (males and females). Here, unreliability was measured as the difference between the 80th and the 50th percentile. It can be questioned whether the RP or SP/RP values will be representative for California as a whole. Many travellers in the SR 91 corridor (Orange County) are quite affluent. The values are certainly not representative for the USA as a whole.

**Brazil**
Senna (1991) derives a model with a mean and a variance of travel time explicitly from utility theory, and arrived at a non-linear utility function, which differs from the state-of-practice, where linear approximations are most popular. In a SP survey in Porto Alegre (Brasil), the results suggest that the valuation of the standard deviation of travel is considerably more important (per minute) than the average travel time (also per minute).

Observations and recommendations

This review has highlighted the existence of a growing body of work on the modelling of the valuation of travel time variability. It should be acknowledged that this review is by no means complete. This is not helped by the fact that some of the existing work is not available in English, or is indeed not publicly available at all. As such, it can for example be assumed that there is a substantial body of work in the private sector, for clients such as railway companies, a point alluded to in König (2004).

For those studies included in our review, some present respondents directly with attributes on the probability of delay (of a certain size), together with travel time and costs. In other surveys, each choice alternative contains a range of possible journey durations (possibly represented graphically), as well as average journey time and costs.

Independent of the approach used, studies in this area typically rely on the use of SP data, a decision that is supported by Bates et al. (2001) and also Noland and Polak (2002). The estimation of a model that includes a reliability variable on RP data is only possible in exceptional circumstances, given issues of correlation between attributes and information about unchosen alternatives.

While all reviewed studies agree that reliability is a factor of substantial importance, there are no generally accepted monetary values for reliability, or indeed a reliable estimate of the relative weight of travel time and travel time reliability. While the majority of studies find that travel time variability is valued more highly than travel time itself, there is some evidence to suggest the opposite, especially in studies that face respondents directly with a measure of reliability. Here, it is possible that some of these observations can be explained on the basis of poor understanding by respondents. Independently of this observation, it should be acknowledged that the valuations of reliability in the present literature come from very specific investigations and are not even used in cost-benefit analyses in the respective countries of origin. Often, the studies are relatively small scale, and in some cases, variability is not the main topic of investigation. The recommendation from this piece of work would thus be that there is a need for a major new study into the valuations of reliability. Such a study would have to investigate both the question of how information on reliability is presented to respondents and how it is later specified in our models. The aim should be to achieve as high a level of consistency between the data and the models as possible, without unduly affecting understanding by respondents or the value of model outcomes. The a priori assumptions as to how respondents evaluate the information in the surveys should be kept to a
minimum. Such work should also look further into the benefits of advanced modelling techniques such as discussed by Michea & Polak (2006) and Liu & Polak (2007).

While highlighting the need for further work, we also recognise the need to have provisional values of reliability that can be used in CBA until new empirical values for the UK become available. Given the lack of representative and generally accepted values, recommendations need to be based on expert opinions in which these experts do their own weighting of the available empirical evidence from various countries, combined with more theoretical/intuitive arguments.

A recommendation is therefore brought forward at this stage for provisional values based on those from the expert workshop organised in Holland (Hamer et al, 2005). In the draft guidance, DfT also used these values for car, although not for public transport (since DfT uses average lateness for this, not RR).

The Workshop ‘Value of Reliability’ took place on 25 October 2004. It was an initiative of the AVV Transport Research Centre of the Dutch Ministry of Transport, Public Works and Water Management, and it was organised by RAND Europe. The aim of the workshop was to provide reasonable provisional values of reliability (VoR) for a range of modes or mode-purpose combinations to be used in CBA in The Netherlands.

A key concept here is that of the reliability ratio (RR). This is defined as the value of reliability of travel time divided by the value of travel time itself. Here we measure reliability as the standard deviation of travel time.

The VoRs presented below are based on the opinions of the experts and the discussions during the workshop. We stress that the VoRs are provisional values. To get evidence-based monetary values for reliability we recommend setting up nationally representative stated preference surveys among car drivers, public transport users, carriers and shippers. For passenger transport by car, the experts agreed on the following reliability ratios (based on the available international evidence, especially from the UK):

**Table II.1: Reliability Ratios for passenger transport by car**

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Reliability Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commuting</td>
<td>0.8</td>
</tr>
<tr>
<td>Business</td>
<td>0.8</td>
</tr>
<tr>
<td>Other</td>
<td>0.8</td>
</tr>
</tbody>
</table>

For public transport, consensus was reached among the experts on the following reliability ratios (based on evidence from the UK, Sweden and The Netherlands):

**Table II.2: Reliability Ratios for public transport**

<table>
<thead>
<tr>
<th>Transport mode</th>
<th>Reliability Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train (interurban)</td>
<td>1.4</td>
</tr>
<tr>
<td>Bus/tram/metro (urban)</td>
<td>1.4</td>
</tr>
</tbody>
</table>
For public transport we recommend to use the same concepts as for car travel: a two-sided (but not necessarily symmetrical) travel time distribution, characterised by a mean (which can be the time according to the timetable) and a standard deviation.

For freight transport, reliability ratios by mode were derived by converting results from earlier Dutch research on the value of time and reliability in freight transport into an RR. For freight transport by road, without further information, we recommend to use an RR of 1.2

All the values mentioned are only provisional, and should not come in the way of new empirical research.

In closing, a point needs to be made concerning the use of the value of reliability. In order to appraise the reliability effects of infrastructure projects in CBAs further work is also needed on the traffic forecasting tools. These need to be improved, so that they are able to provide estimates of the standard deviations or percentiles of travel times on links. Current models typically don’t have this capability. A first pragmatic attempt to develop such a tool in connection to the Dutch National Model system is currently underway in the Netherlands, but it is anticipated that significant further work will be required here.
Annex III: Arup Transport planning Appendix B
APPENDIX C

C1.1.1 Demand Response to Uncertainty and the Weight Given to Major Incidents
C2. Introduction

C2.1 It is helpful to rehearse some of the key conclusions of the theory of departure time choice. A more general review is given in Bates et al (2001) Section 3.3.

C2.2 The simplest form of a “schedule disutility” function is

\[ U(t) = \alpha \Theta(t) + \beta \text{SDE} + \gamma \text{SDL} + \delta D_l \]

where \( \Theta(t) \) is travel time, potentially dependent on the departure time \( t \)
SDE is early schedule delay defined as \( \max(0, \text{PAT} - (t + \Theta(t))) \)
SDL is late schedule delay defined as \( \max(0, (t + \Theta(t)) - \text{PAT}) \)
\( D_l \) is a dummy variable which is equal to 1 if \( \text{SDL} > 0 \) or 0 (otherwise)
\( \alpha, \beta, \gamma \) and \( \delta \) are model parameters, assumed to be negative.

C2.3 Additional disutility is incurred by a discrepancy between the actual arrival time (equal to \( t + \Theta(t) \)) and the “Preferred Arrival Time” (PAT). In most applications, \( \delta \) is implicitly set to zero. It is also often suggested that there is an “indifference band” around PAT, whereby small deviations do not convey any disutility. It is straightforward to redefine SDE and SDL to allow for this.

C2.4 If now we introduce uncertainty in travel time and postulate a distribution of travel time \( f(\Theta(t)) \), say, we can propose that the departure time is chosen by maximising the expected value of the utility (MEU). In the face of uncertain journey times, all the variables in the utility function become random.

C2.5 If a traveller is indifferent as to his arrival time (ie, has no preferred arrival time), then the “schedule” coefficients \( \beta, \gamma \) and \( \delta \) will all be zero, and the choice of departure time will be driven solely by possible variations in the mean travel time: correspondingly, there is no additional disutility due to variability in this case. However, when at least one of the schedule coefficients is non-zero, the general result is for disutility to increase in line with the level of variability. In principle, this additional disutility can and should form a component of “generalised cost” and enter into higher level choice models (eg between alternative modes and destinations).

C2.6 Most authors using this general approach have made the simplifying assumptions that a) \( \delta = 0 \), implying that there is no disutility in being late per se, and b) the anticipated travel time (due to recurrent congestion) is independent of the departure time. In this case, the optimum probability of arriving late is given by \( p^{*}_l = \beta / (\beta + \gamma) \), implying that people choose a departure time to balance the consequences of early and late arrival. This widely reported result is independent of specific assumptions about the distribution of travel time. With some effort it can be shown how it needs to be modified in the face of more realistic assumptions.

C2.7 To find the optimum departure time \( t^* \) we do need to make use of the distribution \( f(\Theta(t)) \), effectively to find the point on the distribution at which \( p^{*}_l = \beta / (\beta + \gamma) \). This then allows us to calculate the optimum value of the expected utility \( EU^* \). In addition to the contribution of the mean travel time \( \alpha \overline{\Theta}(t^*) \), we have the term \( \beta E[\text{SDE}(t^*)] + \gamma E[\text{SDL}(t^*)] \) which can be directly identified with the additional disutility due to JTV.
C2.8 A number of authors have shown that this term is well approximated by \( H(\beta, \gamma) \sigma \) for a wide range of distributions, where \( \sigma \) is the standard deviation of journey time, and \( H \) can be considered constant for any given combination of \( \beta \) and \( \gamma \). This provides some justification for the widespread use of \( \sigma \) as the relevant component in the utility function to indicate the effect of travel time variability.

C3. Sensitivity to the form of distribution

C3.1 It is important to note, however, that the specific value of \( H \) does depend on the form of the distribution \( f(\Theta(t)) \). For example, in the case of a uniform distribution, it can be shown that

\[
H = \sqrt{3} \beta \gamma / (\beta + \gamma)
\]

while for the exponential distribution, Noland & Small (1995) show that

\[
H = \beta \ln \left( 1 + \frac{\gamma}{\beta} \right)
\]

and Black (MVA, 1995) has shown that for the logistic distribution (which is close to the normal)

\[
H = \frac{\sqrt{3}}{\pi} \left( (\beta + \gamma) \ln(\beta + \gamma) - \gamma \ln(\gamma) - \beta \ln(\beta) \right)
\]

C3.2 Dividing \( H \) by the coefficient on mean travel time \( \alpha \) gives what is widely termed the “reliability ratio” [RR]. In general, it can be seen that RR is an increasing function of the ratio \( \gamma / \beta \), so that increases in the relative cost of late arrival leads to reduced optimum utility, essentially because a greater safety margin is required.

C3.3 Note that for all these distributions if either \( \beta = 0 \) or \( \gamma = 0 \) then \( H(\beta, \gamma) = 0 \). Hence, under this formulation, the generalised cost of uncertainty (and as a result the behavioural significance of uncertainty and variability) derives entirely from scheduling considerations.

C3.4 If we make an assumption about the value of \( \beta / \alpha \) (the ratio of early time to travel time), we can calculate the implied value of RR for different distributions using the formula just given. Assuming a “typical” value of 0.5 [note that for theoretical reasons \( \beta / \alpha < 1 \)], we get the picture for three distributions – uniform, logistic and exponential - shown in Figure B.1.

C3.5 This demonstrates some important aspects of the problem. In the first place, for a given set of “scheduling coefficients”, the values of RR are clearly different according to the distribution \( f(\Theta(t)) \). This is driven largely by the “tail” of the distribution. The uniform distribution has no tail, while the logistic is a symmetric distribution, thus with two tails. In practice, however, we expect the travel time distribution not to be symmetric, but positively skewed: the exponential distribution is the simplest form of such a distribution.

C3.6 With low values of \( \gamma / \beta \), implying no marked aversion to lateness, the tail of the exponential distribution is less important, and the impact of the standard deviation is not very different from that for the uniform and logistic distributions. At high values,
however, the relatively unlikely occurrences of high lateness need to be taken into account, and the value of RR is significantly higher for the exponential distribution.

Note that in practice, a value of 2 for the ratio $\gamma/\beta$ is a sensible minimum, since, together with a value of 0.5 for the ratio $\beta/\alpha$, it implies that travellers are no more averse to lateness than to additional travel time. Commonly assumed values for the parameters $\alpha$, $\beta$ and $\gamma$ are 2, 1, 4: the most quoted result, due to Small (1982), gives values corresponding to 2, 1.23, 4.79. This suggests values of $\gamma/\beta$ of the order of 4.

![Figure B.1: Implied “Reliability Ratio”, assuming $\beta/\alpha = 0.5$ (Uniform, Logistic and Exponential distributions)](image)

C3.8 We can further illustrate this by considering an extreme “point” distribution where with a high probability (say $p = 0.99$) delay is zero, but there is a small probability of a large delay $D$ (eg 60 minutes). The standard deviation of such a distribution is readily calculated as $\sqrt{(1 - p) - (1 - p)^2} D$. With such a distribution, it can be shown that there are only two rational choices for the choice of departure time. Either the traveller decides never to be early, in which case he arrives on time in 99% of cases, and accepts the delay of $D$ in 1%, so that his value of $EU = \alpha \tilde{D}(t^*) + (1-p) \gamma D$, or he decides never to be late, in which case he arrives $D$ minutes early in 99% of cases, and on time in 1%, so that his value of $EU = \alpha \tilde{D}(t^*) + p \beta D$.

C3.9 Which of these two options he chooses depends on the relative sizes of $(1-p)\gamma$ and $p \beta$. As long as $\gamma/\beta < p/(1-p)$, which will normally be the case, the traveller will **not** make allowance for the extreme event, and merely accept the (infrequent) disutility of late arrival. However, if there are very severe penalties for late arrival, the traveller will take into account the possibility of extreme events.

C3.10 Note that in this example as well, the disutility due to JTV is proportional to the standard deviation. Compared with the graphs for the continuous distributions, the pattern of the reliability ratio is extreme. If we assume $p = 0.99$, and set $\beta/\alpha = 0.5$ as
before, then as long as $\gamma/\beta < 99$, we will have $RR = 0.5(1-p)(\gamma/\beta)/$
$\sqrt{(1 - p) - (1 - p)^2} = .0503 \, \gamma/\beta$, thus ranging from 0.1 when $\gamma/\beta = 2$ up to 4.97 when $\gamma/\beta = 99$. Thereafter it remains constant.
C3.11 If then we re-draw Figure B.1 adding the “extreme” distribution, we obtain Figure B.2.

![Graph showing implied Reliability Ratio (RR) vs. γ/β for different distributions]

**Figure B.2: Implied “Reliability Ratio”, assuming β/α = 0.5**

(Uniform, Logistic, Exponential and “Extreme” distributions)

C4. Although these results are entirely based on theory, they do cast some important light on the treatment of “extreme” events. By their nature, extreme events are related to incidents as opposed to other sources of JTV. The theoretical discussion above suggests that, firstly, whether extreme events are discounted is dependent on the lateness penalty. If this is not particularly high, then the theory predicts that people will behave as if they discount extreme events.

C4.1 On the other hand, the discussion makes clear how the reliability ratio is dependent on the form of the distribution. Assuming a value of γ/β between 3 and 4, consistent with empirical evidence, RR is around 0.65 for the uniform and logistic distributions, varies between 0.7 and 0.8 for the exponential, and between 0.15 and 0.2 for the extreme (assuming a 1 in 100 chance of an extreme event). It is noteworthy that the value currently assumed is 0.8 (based on the Cranfield 1993 work) for all vehicles except OGVs, where a value of 1.2 is used. This is broadly compatible with the exponential distribution. However, if the true distribution was of the “extreme” variety, the value should be deflated by a factor of between 4 and 5.

C4.2 In practice, as noted, we do not expect the distributions to be symmetric. Moreover, while in terms of variations in the travel time distribution due to random demand and capacity variations, we might expect the shape of the travel time distribution to stay roughly the same, when it comes to major incidents on the road network it is quite clear that they change the shape of the travel time distribution dramatically. It seems important, therefore, to explore the consequences over a fairly typical range of values. To do this a simple simulation model can be constructed.
C5. Comparing utility estimates from simulation

C5.1 Purely for convenience we assume a Normal distribution for travel time with an expected value of 60 minutes and a standard deviation of 6 minutes. The disutility of travel, early and late time are defined as 2, 14, in line with earlier discussion. Under these assumptions the optimum probability of lateness [β/(β+γ)] is 0.2, and this is achieved with a departure time of 65 minutes before PAT. In other words, the probability of the travel time being less than 65 minutes is 0.8.

C5.2 With this optimum departure time, the distribution of arrival time is shown in the upper dotted curve in Figure B3. The various possible arrival times give rise to different amounts of early or late schedule delay. By applying the values to these amounts and allowing for the distribution of actual arrival times we can calculate the mean schedule disutility, over and above the travel time disutility. In units of travel time, this turns out to be 4.19 minutes, and since the standard deviation is 6 minutes, gives an RR value of 0.70. Note that in utility terms the coefficient on sigma is double this (ie 1.40), since we have assumed a value of 2 for travel time.

![Image](image_url)

**Figure B.3: Travel Time Distribution with and without Major Incidents**

C5.3 We now consider the possibility of major incidents. Suppose there is the possibility of a major incident leading to additional delay according to a uniform distribution between 0 and 60 minutes. When the probability of this major incident is p=0.2 the new distribution of arrival times is shown by the continuous curve in Figure B.3; we have assumed that there is no change in departure time.

C5.4 With the information about the distribution of travel time we can calculate the utility loss as a result of the extra lateness caused by an incident. There are two elements – the loss due to increased travel time and the loss associated with the increase in variability. The latter can be estimated in two ways:

**Method 1**

C5.5 We can find the new standard deviation of the distribution and hence the increase in the standard deviation over the base (TAMTTV and Cohen & Southworth adopt this method). Then, using the observed utility per standard deviation (1.40 in the
example), we can calculate the loss of utility \( \Delta U = \beta_\sigma (\sigma_2 - \sigma_1) \), using subscripts 1 and 2 to represent before and after.

**Method 2**

**C5.6** Alternatively, we can calculate the reduction in expected utility due to the extra late time (and the small reduction in early time):

\[
\Delta U = \beta (E[SDE_1] - E[SDE_2]) + \gamma (E[SDL_2] - E[SDL_1])
\]

**C5.7** In both these calculations we assume initially that the departure time decision is unchanged; in other words travellers do not change their departure time in response to the new distribution.

**C5.8** The outcome of these two calculations can be seen in Table B.1. The first line of results reproduces the base position before incidents: the total average loss of 8.379 per traveller, divided by the standard deviation, gives the multiplier 1.40.

**Table B.1**  
**Average Utility Calculations By The Two Methods**

<table>
<thead>
<tr>
<th>Probability of Incident Occurring (p(%))</th>
<th>Standard Deviation of Journey Time ((\sigma))</th>
<th>Disutility from earliness (1)</th>
<th>Disutility from lateness (4)</th>
<th>Disutility from earliness &amp; lateness</th>
<th>Change in disutility (Method 2)</th>
<th>Change in disutility (Method 1)</th>
<th>Ratio of the two disutility estimates (Method 1/Method 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>6</td>
<td>5.706</td>
<td>2.673</td>
<td>8.379</td>
<td>0.939</td>
<td>1.287</td>
<td>1.37</td>
</tr>
<tr>
<td>1.0</td>
<td>6.9</td>
<td>5.654</td>
<td>3.664</td>
<td>9.318</td>
<td>0.939</td>
<td>1.287</td>
<td>1.37</td>
</tr>
<tr>
<td>2.5</td>
<td>8.1</td>
<td>5.576</td>
<td>5.150</td>
<td>10.726</td>
<td>2.347</td>
<td>2.918</td>
<td>1.24</td>
</tr>
<tr>
<td>5.0</td>
<td>9.7</td>
<td>5.445</td>
<td>7.628</td>
<td>13.073</td>
<td>4.693</td>
<td>5.143</td>
<td>1.10</td>
</tr>
<tr>
<td>10.0</td>
<td>12.1</td>
<td>5.184</td>
<td>12.582</td>
<td>17.766</td>
<td>9.386</td>
<td>8.554</td>
<td>0.91</td>
</tr>
</tbody>
</table>

**C5.9** The difference in the results from the two methods is revealing. Table B.1 shows the results for \( p = 0.0\% \) to 10.0\%. The disutility from earliness is the product of the earliness (in minutes) and the value of earliness. (The disutility from lateness is defined similarly). As the proportion of travellers affected by the major incident increases, the standard deviation of travel time predictably increases. Disutility from earliness declines slightly - again as expected. The big impact of course is in lateness. Average disutility here increases from 2.673 to 12.582 for \( p=10\% \) (the scale of this change is close to 10\% of travellers experiencing an extra 30 minutes lateness with a value of 4 per minute).

**C5.10** The ‘changes in disutility’ refer to the estimate of utility loss from the two methods and refer to the change in utility from the base position before incidents (\( p = 0.0\% \)). The last column compares the estimates of utility loss from the ‘earliness/lateness’ method and the ‘standard deviation’ method. Where the probability of an incident is 1\%, the standard deviation method gives a figure 37\% higher than the lateness
method. This figure falls until there is approximate equality when the probability of an incident is 7-8%.

C5.11 The assumption made so far is that travellers do not adjust their departure time in response to their perception of the probability of incidents. If we now assume that they optimise their departure time with full knowledge of this probability then a slightly different picture emerges (Table B.2). As the probability of incidents rises, travellers move their departure time earlier. This leads predictably to more earliness and less lateness. Although as a result the sum of lateness and earliness disutility falls, it is only a slight reduction, reflecting the fact noted in the previous section that as long as the probability of the incident remains low, it will only have a significant influence on departure time if the penalties for late arrival are very high.

Table B.2 Average Utility Calculations By The Two Methods
( Optimised Departure Time;
Incident delay Uniformly distributed between 0 and 60 Minutes)

<table>
<thead>
<tr>
<th>Probability of Incident Occurring (p(%))</th>
<th>Standard Deviation of Journey Time (\sigma)</th>
<th>Disutility from earliness [\beta = 1]</th>
<th>Disutility from lateness C5.11.1.1 [\gamma = 4]</th>
<th>Change in disutility from earliness &amp; lateness</th>
<th>Change in disutility (Method 2)</th>
<th>Change in disutility (Method 1)</th>
<th>Ratio of the two disutility estimates (Method 1/Method 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>6</td>
<td>5.706</td>
<td>2.673</td>
<td>8.379</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>6.9</td>
<td>5.776</td>
<td>3.539</td>
<td>9.315</td>
<td>0.936</td>
<td>1.287</td>
<td>1.38</td>
</tr>
<tr>
<td>2.5</td>
<td>8.1</td>
<td>5.888</td>
<td>4.821</td>
<td>10.709</td>
<td>2.329</td>
<td>2.917</td>
<td>1.25</td>
</tr>
<tr>
<td>5.0</td>
<td>9.7</td>
<td>6.094</td>
<td>6.907</td>
<td>13.001</td>
<td>4.622</td>
<td>5.143</td>
<td>1.11</td>
</tr>
<tr>
<td>10.0</td>
<td>12.1</td>
<td>6.607</td>
<td>10.848</td>
<td>17.455</td>
<td>9.075</td>
<td>8.554</td>
<td>0.94</td>
</tr>
</tbody>
</table>

C5.12 The comparison of the two methods for calculating utility remains similar to the case where optimisation was based on zero expectation of an incident - a large difference for low p with near equality when p is between 5 and 10%.

C5.13 Another factor that can affect the difference between the utility estimates from the two methods is the relative value of earliness and lateness (\gamma/\beta). Figure B.4 shows that the standard deviation method tends to overestimate the benefits at low values of \gamma/\beta (around 2.0), but to underestimate for higher values of \gamma/\beta (around 8.0). The departure time is assumed to be unchanged. The different curves reflect different assumptions about the probability of a major incident: the ratio of the two utility estimates falls as the probability increases (i.e., the occurrence of an incident becomes “less extreme”). Hence there are some combinations of the ratio \gamma/\beta and the probability of a major incident where the two estimates coincide, but this is more or less fortuitous.

C5.14 Finally we examine the case (Figure B.5) where the incident is more severe with a uniform distribution covering the range 0-120 minutes rather than 0-60 minutes. In the case of a more severe incident the difference between the two methods is greater
with the standard deviation method giving a result almost always higher and in some cases (with plausible lateness and incident probability parameters) two times higher.

From this collection of results it is clear that the standard deviation method usually gives quite different results from the method using values of earliness and lateness directly. The ratio of the results varies but can be as large as two for quite plausible assumptions. In general the divergence between the two methods becomes larger as the duration of the incident increases.

**Figure B.4** Ratio of Utility Estimates from the Two Methods
(Value of $\gamma/\beta$ varies between 2 and 8;
Probability of a major incident varies between 0.5% and 10%;
Incident delay Uniformly distributed between 0 and 60 minutes)

**Figure B.5** Ratio of Utility Estimates from the Two Methods
(Value of $\gamma/\beta$ varies between 2 and 8; Probability of a major incident varies between 0.5% and 10%; Incident delay Uniformly distributed between 0 and 120 minutes)
Annex IV: References


Department for Transport (2004a) TAG Unit 2.5 The Appraisal Process. Online at: www.webtag.org.uk

Department for Transport (2004b) TAG Unit 3.5.1 The Public Accounts Sub-Objective. Online at: www.webtag.org.uk

Department for Transport (2004c), TAG Unit 3.5.2 The Transport Economic Efficiency Sub-Objective. Online at: www.webtag.org.uk.


Department for Transport (2005), Transport, Wider Economic Benefits and Impacts on GDP. London: DfT.

Department for Transport (2006a) TAG Unit 3.5.4 Cost-Benefit Analysis. Online at: www.webtag.org.uk.

Department for Transport (2006b) TAG Unit 3.5.9 The Estimation and Treatment of Scheme Costs. Online at: www.webtag.org.uk.

Department for Transport (2007a) TAG Unit 3.5.6 Values of Time and Vehicle Operating Costs. Online at: www.webtag.org.uk.

Department for Transport (2007b) TAG Unit 3.5.7 The Reliability Sub-Objective, unpublished draft, March 2007.

101


MVA (1996) Benefits of reduced travel time variability; report to DfT; MVA.


Transek (2001) Stockholm public transport study; Transek AB, Stockholm


