

***Understanding the Mean-Variance framework  
through the application of Public Transport  
Smartcard data***

Christopher John Leahy

Submitted in accordance with the requirements for the degree of  
Doctor of Philosophy

The University of Leeds  
Institute for Transport Studies (ITS)

May 2017



### *Intellectual property and publication Statements*

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

### *Assertion of moral rights*

The right of Christopher John Leahy to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

©2017 The University of Leeds and Christopher John Leahy

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The literature review in Chapter 2 of the thesis on ‘Reliability Equivalence’ has appeared in publication as follows:

LEAHY, C., CHEN, H. & BATLEY, R. 2014. Reliability Equivalence in Public Transport Contexts. *Procedia - Social and Behavioral Sciences*, 138, 185-192.

I was responsible for the literature review and application of reliability equivalence using smartcard data. The contribution of the co-authors was providing advice on the data and guidance on the economic theory of the paper.

The work in Chapter 5 of the thesis has appeared in publication as follows:

LEAHY, C., BATLEY, R. & CHEN, H., 2016. Toward an automated methodology for the valuation of reliability. *Journal of Intelligent Transportation Systems*, 20(4), 334-344.

I was responsible for the contextualisation of the problem, data analysis model application and interpretation. The contribution of the co-authors was advising on the application discrete choice models and their interpretation.

The work in Chapter 6.2 of the thesis has appeared in similar form in publication as follows:

LEAHY, C. 2016. The Development of an Improved Microeconomic Framework for Transportation Reliability. *The International Journal of Transport Economics*, 43(1-2), 11-31.

I was responsible for the contents this paper. However my supervisors are acknowledged for their comments and encouragement in its development

## **Acknowledgements**

My thanks are due to my supervisors, Prof Richard Batley and Dr Haibo Chen who have guided through the PhD process and provided valuable insights and lessons along the way.

I would also like to thank Transport for London (TfL) who have supported this project financially and provided the datasets and advice on its direction. Particular thanks are due to Maeve Clements and Duncan Horne with respect to the data.

Finally, my thanks are due to my partner Amy, as well as my family and friends who have provided encouragement and support throughout the period of study.

## **Abstract**

Reliability is an important aspect of any transportation system, and there has been a substantial amount of research to bring it to the point where it can be accurately incorporated into established forecasting and appraisal frameworks. At the same time, emerging data sources, such as public transport smartcards, provide opportunities to understand more about the reliability of a particular transport system.

This thesis conducts research on reliability using smartcard data. In the first instance the thesis provides a critique of the Mean-Variance framework for the treatment of transport reliability and finds room for adaptation. The thesis also provides a review of empirical research that estimates Mean-Variance variables and parameters, and finds evidence of methodological issues.

In response to these issues, the thesis utilises smartcard data to investigate Mean-Variance in three ways. The key element is the development of an alternative method of estimating a value of reliability, treating smartcard data as a Revealed Preference data source and combining it with established discrete choice methods. The second element uses the smartcard data to identify the factors affecting reliability levels through estimation of a linear regression model. The third strand of investigation employs the data to understand more about possible alternative measures of reliability and compares the underlying utility function of Mean-Variance with a second reliability framework.

The thesis therefore demonstrates that the application of public transport smartcard data has the potential to yield insights in the field of transport reliability. In particular, it establishes how this data source might be used to estimate the value of reliability. With development, it may have the potential to forecast future reliability levels. Application of the data also supports the status quo of utilising the standard deviation as an indicator of reliability.

## Contents

<b>CHAPTER 1 - INTRODUCTION</b> .....	<b>1</b>
<b>1.0 INTRODUCTION</b> .....	<b>1</b>
1.0.1 <i>Reliability Definition</i> .....	1
1.0.2 <i>Reliability and Risk</i> .....	2
1.0.3 <i>The Standard Approaches to Reliability</i> .....	3
1.0.4 <i>Quantification of the ‘Value of Reliability’</i> .....	4
1.0.5 <i>New Data Sources</i> .....	4
1.0.6 <i>Smartcard Analysis in Public Transportation</i> .....	6
1.1 MOTIVATION.....	6
1.1.1 <i>Policy Context</i> .....	7
1.2 AIMS AND OBJECTIVES.....	8
1.3 NOVELTY AND ORIGINAL CONTRIBUTION.....	12
1.4 THESIS OUTLINE.....	13
<b>CHAPTER 2 - LITERATURE REVIEW</b> .....	<b>15</b>
2.0 INTRODUCTION.....	15
2.1 EXPECTED UTILITY THEORY.....	17
2.2 MEAN-VARIANCE: BACKGROUND AND DEVELOPMENT.....	23
2.2.1 <i>Portfolio Theory: Early Development</i> .....	23
2.2.2 <i>Mean-Variance and Transportation Reliability</i> .....	28
2.3 <i>Mean-Variance Indicators</i> .....	33
2.2.4 <i>Conclusion</i> .....	35
2.3 ALTERNATIVE RELIABILITY FRAMEWORKS.....	36
2.3.1 <i>The Scheduling Approach</i> .....	36
2.3.2 <i>The Mean-Lateness Approach</i> .....	42
2.3.3 <i>Reliability Equivalence</i> .....	43
2.4 EMPIRICAL APPLICATION OF RELIABILITY FRAMEWORKS.....	46
2.4.1 <i>Stated Preference</i> .....	46
2.4.2 <i>Revealed Preference</i> .....	58
2.4.3 <i>Summary of Studies</i> .....	63

2.4.4 Empirical Support for Equivalence.....	76
2.5 SUMMARY AND CONCLUSIONS .....	78
<b>CHAPTER 3 - DATASETS.....</b>	<b>83</b>
3.0 INTRODUCTION.....	83
3.1 BACKGROUND TO SMARTCARD TICKETING.....	83
3.1.1 Introduction.....	83
3.1.2 Literature Review .....	85
3.2 DATASETS .....	89
3.2.1 TfL DATASETS.....	89
3.3 DATASET ISSUES .....	93
3.3.1 Pre-analysis Issues.....	94
3.3.2 Post-analysis Issues .....	100
3.4 INITIAL RELIABILITY ANALYSIS .....	102
3.4.1 Travel Time Distribution.....	102
3.4.2 Analysis of Reliability Indicators .....	105
3.5 CONCLUSION .....	118
<b>CHAPTER 4 - THE FACTORS AFFECTING RELIABILITY ON THE LONDON UNDERGROUND .....</b>	<b>122</b>
4.1 INTRODUCTION.....	122
4.2 PREVIOUS WORK.....	122
4.2.1 Forecasting Reliability for Highway Modes.....	122
4.2.2 Forecasting Reliability for Public Transport.....	124
4.3 MODEL VARIABLES .....	130
4.4 DATA.....	133
4.5 MODEL ESTIMATION .....	137
4.6 DISCUSSION.....	139
4.7 CONCLUSION .....	140
<b>CHAPTER 5 - DISCRETE CHOICE MODELLING WITH SMARTCARD DATASETS.....</b>	<b>142</b>
5.0 INTRODUCTION.....	142

5.1 BACKGROUND .....	143
5.1.1 <i>Travel Time Variation (TTV)</i> .....	143
5.1.2 <i>The Revealed Preference approach</i> .....	145
5.2 MODELLING FRAMEWORK .....	146
5.2.1 <i>Choice and Risk</i> .....	146
5.2.2 <i>Expected Utility</i> .....	146
5.2.3 <i>Data Requirements</i> .....	147
5.2.4 <i>Choice Model Specification</i> .....	147
5.3 EMPIRICAL APPLICATION OF THE MODEL.....	150
5.3.1 <i>Smartcard Data</i> .....	150
5.3.2 <i>Data Analysis</i> .....	153
5.4 CHOICE MODELLING .....	161
5.5 DISCUSSION.....	167
5.6 CONCLUSION .....	169
<b>CHAPTER 6 – ALTERNATIVES TO MEAN-VARIANCE.....</b>	<b>172</b>
6.1 INTRODUCTION.....	172
6.2 ALTERNATIVES TO VARIANCE IN MV .....	172
6.3 MEAN-VARIANCE AND SCHEDULING UTILITY FUNCTIONS .....	178
6.3.1 <i>Issues With Matching MV and Scheduling Utility Functions</i> .....	184
6.3.2 <i>The Reliability Premium</i> .....	186
6.4 CONCLUSION .....	189
<b>CHAPTER 7– CONCLUSION .....</b>	<b>193</b>
7.1 INTRODUCTION.....	193
7.2 SUMMARY OF THE THESIS .....	193
7.3 AIMS AND OBJECTIVES.....	195
7.4 LIMITATIONS OF THE STUDY .....	199
7.5 FURTHER WORK ARISING FROM THIS STUDY .....	201
7.6 SUMMARY AND CONTRIBUTION .....	202
<b>REFERENCES.....</b>	<b>204</b>



Appendix 1 - Scatter plots of statistically significant explanatory variables against standard deviation of journey time from Table 4.5

Appendix 2 – Variables used to estimate discrete choice models in Chapter 5

Appendix 3 – XY plots of relationships between explanatory variables

## List of Figures

Figure 2.1 - Illustration of a concave (risk averse) utility function and featuring a single risky prospect with two outcomes .....	20
Figure 2.2 - Illustration of a concave (risk averse) utility function and two risky travel time prospects .....	21
Figure 2.3 - Expected payoff and standard deviation combinations with an efficient frontier line (reproduced from Markowitz, 1959).....	27
Figure 2.4 - A Scheduling utility function based upon Small (1982).....	39
Figure 2.5 - A Scheduling utility function based upon Small (1982) and Batley (2007).....	40
Figure 3.1 - Average number of entries and exits by ticket type and day (Edgware to Camden Town inclusive) .....	99
Figure 3.2 - Profile of weekday use by 15 minutes, normalised between Oyster and magnetic tickets.....	100
Figure 3.3 - Percentiles of travel time from Edgware to Camden Town by 15 minute time segment .....	103
Figure 3.4 - Empirical probability density function for all weekday travel time data Edgware to Camden Town.....	104
Figure 3.5 - Empirical probability density functions for AM Peak (AMP), PM Peak (PMP) and Inter peak (INT) periods, Edgware to Camden Town .....	105
Figure 3.6 - Chart of reliability indicator levels (ALL = all data, AMP = AM peak data, INT = inter peak data, PMP = PM peak data).....	111
Figure 3.7 - Cumulative distribution of indicators used for within day analysis....	114
Figure 3.8 - Probabilistic reliability indicator for each time period, varying by $\lambda$ parameter.....	115
Figure 5.1 - A mapping of rail based modes to composite modes used in subsequent modelling .....	153
Figure 5.2 - Mean journey time difference plotted against mode choice .....	156
Figure 5.3 - Standard deviation of journey time difference plotted against mode choice .....	156
Figure 5.4 - Headway differences plotted against mode share .....	157

Figure 5.5 - Scatterplot of mean travel time against standard deviation of travel time for 62 OD/mode combinations.....	158
Figure 5.6 - OD plot of the standard deviation of the Metro/Heavy Rail mode against the proportion using the Metro/Heavy Rail mode .....	164
Figure 6.1 - An example MV utility function .....	174
Figure 6.2 – MV utility function (travel time risk after PAT only) .....	175
Figure 6.3 - Plot of Scheduling utility function based upon typical parameters where $\beta > \alpha > \gamma$ (estimated by Li et al, 2010) and omitting a fixed penalty for lateness ( $\theta = 0$ ) .....	178
Figure 6.4 - Plot of equivalent utility functions, where a quadratic function is fitted to a Scheduling function based upon overall EU and average slope of the Scheduling function .....	181
Figure 6.5 - A two prospect example demonstrating the possibility of equivalence between Scheduling and MV utility functions based upon two conditions of equivalence .....	182
Figure 6.6 - A Scheduling utility function, with the risk angle, $\Theta$ .....	183
Figure 6.7 - A quadratic utility function fitted to a simple Scheduling utility function, based upon the three conditions.....	184
Figure 6.8 - A quadratic utility function fitted to four Scheduling utility functions, based upon the three conditions of equivalence.....	185
Figure 6.9 - A concave utility function with two risky travel time prospects, the expected value of t, E(t), and the certainty equivalent, $t_c$ .....	187

## List of Tables

Table 1.1 - Full list of objectives to be met in this thesis .....	10
Table 2.1 - Implied ML elasticities derived from SP estimates of the value of Mean-Lateness.....	56
Table 2.2 - Summary of revealed AML Elasticities (from Wardman and Batley, 2014) .....	63
Table 2.3 - Summary table of key reliability studies .....	65
Table 2.4 - Summary statistics of RR studies in Table 2.3.....	70
Table 2.5 - Results of meta-analysis using OLS regression .....	72
Table 2.6 - Objectives met by Chapter 2 .....	79
Table 3.1 - Description of fields within the ‘SEQUENCES_JRNYS’ datasets.....	91
Table 3.2 - Ranked use of stations in Edgware to Camden Town study area by total number of entries and exits, split by Oyster and magnetic trips .....	98
Table 3.3 - Candidate indicators of reliability .....	107
Table 3.4 - Weekday summary statistics for Edgware to Camden Town OD based on Oyster data only .....	109
Table 3.5 - Indicators of reliability for time periods across days .....	110
Table 3.6 - Spearman correlation coefficient between all candidate indicators of reliability .....	113
Table 3.7 - Objectives met by Chapter 3 .....	119
Table 4.1 - Model estimation results from Uniman (2009) .....	128
Table 4.2 - Summary statistics calculated across ODs .....	135
Table 4.3 - Correlation matrix of initial variables .....	136
Table 4.4 - Parameter estimates of Model 1 .....	138
Table 4.5 - Parameter estimates of Model 2 .....	139
Table 5.1 - Summary statistics of dataset used to estimate choice models.....	155
Table 5.2 - Correlation matrix of explanatory variables .....	159
Table 5.3 - Number of uses by Oyster card ID .....	160
Table 5.4 - Parameter estimates from MNL .....	162
Table 5.5 - Parameter estimates from MNL with interaction terms .....	163
Table 5.6 - Mean-Variance marginal utilities and RR by mode .....	163

Table 5.7 - Parameter estimates from MNL with additional variables .....	165
Table 5.8 - Model parameters from CNL.....	166
Table 5.9 - Allocation parameters from CNL .....	167
Table 5.10 - Objectives met in Chapter 5 .....	170
Table 6.1 - Model results for each risk indicator using RP dataset.....	176
Table 6.2 - Reliability premia calculated for four cases of matching a MV utility function to a Scheduling utility function. All values presented in units of travel time, t. .....	188
Table 6.3 - Objectives met by Chapter 6.....	191

## List of Abbreviations

AMP	AM peak
APC	Automatic payment collection
AVL	Automatic vehicle location
CNL	Cross nested logit
DLR	Docklands light rail
EJT	Excess journey time
EU	Expected utility
HOT	High occupancy toll (lane)
INT	Inter peak
LU	London Underground
MNL	Multinomial logit
MV	Mean-variance
OD	Origin-destination
PMP	PM peak
PT	Public transport
RP	Revealed preference
SP	Stated preference
RR	Reliability Ratio
TTV	Travel time variation
U	Utility
VOR	Value of reliability
VOT	Value of time

# Chapter 1 - Introduction

## 1.0 Introduction

The first chapter in this thesis will introduce the two topics that will form the focus of subsequent work: the *reliability* of transport and the use of *smartcard* data in research contexts.

The chapter begins by providing some initial background to these topics. Subsequently the motivation for the study and the policy context of reliability is examined in more detail; with a focus upon the UK and specifically Transport for London (TfL) who are supporting this project.

The aims and objectives of this study will then be identified based upon the background to the topics. The next section will outline the contribution that this thesis makes to the field of transportation reliability, and finally an outline of the work contained within the remainder of the thesis will be provided.

### 1.0.1 Reliability Definition

The term *reliability* can be broadly used in transportation and is open to a variety of interpretations. It could be applied in relation to the cleanliness or crowding level of a public transport vehicle over time. However in context of this thesis, the term is used exclusively in relation to *travel time*.

To give an example of what is meant by travel time reliability, a commuter may complain about arriving late for work as a result of unusually heavy traffic. To reach a precise definition of reliability, it is important to understand what such a complaint might refer to. The complaint implies that a longer travel time was incurred due to a random element of travel time which made the traveller arrive later than their work start time. It is well established that in a majority of transportation contexts extra travel time is viewed negatively by travellers. However in this case it is not the extra travel time *per se* that is of interest, but rather the negative consequences of unexpected travel time in relation to some subsequent activity.

What this example alludes to is a travel time *expectation* that is not met. In public transport contexts, a traveller's expectations are often clearly set by a published timetable. Newer technologies such as route planning software and satellite navigation systems also help to calibrate a traveller's expectations when using other modes such as the private car. When time-related expectations are not met, the traveller is likely to incur some impact: in the example above, the traveller is late for work and may possibly lose income as a result; he/she might also experience stress related to the *uncertainty* of their travel time.

What is conveyed with this introduction is that when the performance of a transport system is at odds with a traveller's expectations, some impact will be felt. Furthermore, such an impact may affect future transport decisions such as choice of departure time, route or mode, or even whether to travel; responses which give good reason for wanting to understand the *value of reliability* (VOR). The VOR is the weight that a traveller places on a marginal unit of reliability. It is commonly expressed in units of money, or *willingness to pay* to realise an improvement in reliability (or to avoid a deterioration). This concept will be expanded upon in Chapter 2.

### **1.0.2 Reliability and Risk**

The term *travel time risk* is sometimes used in transport economics in place of *reliability*; although they are essentially referring to the same phenomenon. The term *risk* comes with its own definition from the field of economics; referring to a situation where a number of outcomes could arise from an economic decision, and each has a known probability associated to it (Knight, 1921). This would be, for example, the case of a gamble where the participants have to guess the outcome of a fair coin toss. A related term to risk is *uncertainty*, where the probability of some or all of the outcomes occurring are unknown e.g. the return on a financial investment. In transport applications, it is risk which is most commonly used, in combination with *expected utility theory* (EUT) as the underlying microeconomic framework. It is under EUT that the methods for quantifying the VOR have been developed. A fuller exploration of risk and EUT will also take place in Chapter 2.



What the term *travel time risk* therefore implies is that a travel option will have a number of probabilistic travel time outcomes associated with it. It is with this idea in mind that the *travel time distribution* can be introduced; a concept that is statistical in nature. When referring to reliability in statistical terms, it is more appropriate to refer to *travel time variation*. In the work that follows, the terms *reliability*, *travel time risk*, and *travel time variation* are used as appropriate to the discussion taking place.

### **1.0.3 The Standard Approaches to Reliability**

The characterisation of the reliability of a transport trip as related to a statistical distribution is an attractive concept. It is this idea, first developed in relation to risk on financial investments (Markowitz, 1952), which has prompted a measure of the width of a distribution to be used as an indicator of risk. Jackson and Jucker (1982) are attributed with the introduction of this concept into transport contexts. The *standard deviation*, as the most common measure of dispersion, has become the predominant measure of travel time variation. The standard deviation is often used to characterise reliability on a transport trip or link, as well as to estimate a VOR. In the latter context, it is typical to describe a traveller's expected utility of a trip as a function of both travel time and the standard deviation of travel time. This framework is commonly known as the *Mean-Variance* (MV) approach and has been applied extensively in estimating the VOR. The MV approach will form the basis of what is to follow in this thesis and will be explored in depth in the literature review chapter. Subsequently, MV parameters and variables are estimated using established modelling techniques, and an investigation into alternative risk measures will be conducted.

The key alternative method for approaching reliability is commonly referred to as the *Scheduling* approach (Small, 1982; Noland and Small, 1995). It recognises that the key consequence of travel time risk is a probability of the traveller arriving early or late at their destination, both of which will negatively affect the *utility* of the traveller. As such, the traveller will often build extra time into their trip, known as the *headstart*, in order to minimise this impact (Gaver, 1968). The Scheduling approach should not be confused with public transport schedules and timetables. Therefore the approach is capitalised throughout this thesis to indicate this distinction.

The final key framework for the treatment of reliability is *Mean-Lateness*, which is related to the Scheduling approach insofar as it takes account of average lateness at the destination of a trip. It is primarily applied within the UK rail industry, and therefore the lateness is measured in relation to the set timetable.

#### **1.0.4 Quantification of the ‘Value of Reliability’**

The accurate quantification of the VOR has been a primary objective of research utilising the above frameworks. There have been efforts to agree a standard valuation of reliability among experts (e.g. de Jong et al, 2009), but nevertheless ambiguity remains. Meta-analyses in the field (such as that of Carrion and Levinson, 2012; Wardman and Batley, 2014) demonstrate that a broad range of estimates exist within the academic literature. The value of reliability will vary depending on various features of the journey and traveller, including demographics, mode, and the design of the survey itself (Li et al, 2010). The cost of conducting a Stated Preference (SP) survey for the estimation of a VOR means that it is commonplace for a previously estimated VOR to be *transferred* to a new context. It is reasonable to question the appropriateness of this practice when the potential exists for the VOR to vary widely for the reasons given above. Such practical issues can make applications of the VOR difficult. New and emerging data sources based upon automated systems offer the potential to:

- overcome the practical problem of cost in surveying travellers;
- provide a value of reliability suitable to the temporal and geographic context of the application
- reveal real-world behaviour.

#### **1.0.5 New Data Sources**

One of the key innovations in transportation in the recent past has been the introduction of information technology. Transport authorities, operators and consumers have increasingly adopted such systems into their routine operations or behaviour. Examples of these systems include:

- Automatic number plate recognition, as used by operators of cordon congestion charging zones to monitor those drivers entering and making trips

within an area where a charge is levied. These systems are able to detect number plates via roadside cameras and check vehicles against payment records;

- Smart Motorways, utilising traffic cameras and sensors to manage traffic flows on the UK's strategic road network;
- Fleet tracking systems for use by hauliers to more efficiently manage their operations. Such systems rely upon a global positioning system (GPS) in each vehicle of the fleet. The position of the vehicle is communicated back to a central system, allowing the fleet to be monitored, co-ordinated and directed centrally.

Whilst such systems have an immediate benefit to the individuals or organisations that use them, they will also generate large amounts of data. Many developers of such systems have recognised the value of such data and have made use of it themselves to improve their products or provided it to partners. For example, in the case of Smart Motorways, the UK's Highways Agency processes this data to provide travel information to media outlets. In another example, the logistics operator DHL utilises traffic condition data provided by satellite navigation systems on board other road vehicles in order to calculate optimal routing for its own fleet (POST, 2014).

Similar systems have also emerged within public transportation contexts. Automatic payment collection (APC) systems are increasingly common on urban public transport networks (Pelletier, 2011). Such systems usually require a passenger to possess a card capable of holding account data. The card is scanned into the transport system at the beginning of the trip and, where a distance based fare is levied, at the end of the trip also. Contactless or swipe technology is commonly utilised, such that the card is capable of communicating with the fare collection system with minimal delay to the passenger (Blythe, 2004). Cards used in this way are known as *smartcards*, which is a concept and term that will be employed throughout this thesis. These cards are widely used where offered, due to the benefit to passengers of reduced cognitive burden (being released from the calculation of their fare for the trip), lessening the need to carry money and reducing the time taken to pay for their trip. The technology also reduces the burden on the public transport operator or authority by simplifying their cash handling

procedures. A subsidiary benefit of smartcards is that they generate large volumes of data related to passenger behaviour, choices and experience: TfL figures would suggest 21.6M passenger trips are tracked on an average day. Research efforts have previously been made to assist practitioners in making use of this resource, and the present study will contribute to this field by utilising public transport smartcard data in the investigation of reliability.

### **1.0.6 Smartcard Analysis in Public Transportation**

Smartcard data has formed the basis of a small body of research. After obtaining the datasets, a key task for researchers is to understand how to practically analyse them. It is also necessary to appreciate that the data outputs are a by-product of some other process; the data may not always contain the level of accuracy required for the research application at hand. A detailed data description is vital to understanding the possibilities and limitations of such datasets.

Previous applications of smartcard data will be covered in Chapter 3 of this thesis. There are useful examples of methods with which to enrich the data such as linking trips into chains. Most of the analyses of smartcards have been with a view to obtain information on system performance or demand levels separately, but there is little interaction between the two. In the work that follows, use of the smartcard data will be demonstrated that progresses beyond basic analysis. This will be done by exploring the dataset's suitability as an input for the estimation of *linear regression* and *discrete choice models*, in combination with aforementioned MV reliability framework.

## **1.1 Motivation**

This study is motivated by an acknowledgment that reliability remains a relatively peripheral aspect of the appraisal process in comparison to travel time savings, despite some pre-existing studies suggesting that the VOR (the ratio of the marginal utility of standard deviation of travel time to the marginal utility of cost) is in excess of the *value of travel time* (VOT) (Senna, 1994; Liu et al, 2004). The reasons for the less comprehensive treatment of reliability in transport planning are unclear, and therefore this study begins by outlining and critiquing the current state of the art in the field. The literature review will demonstrate an appreciation of the theoretical foundations of

reliability. This discussion will be developed to the key empirical studies related to the VOR and consideration given to whether improvements to the predominant methods used to estimate this value will improve the acceptability of reliability in appraisal.

This introduction has already described how new technologies have improved travel for a range of stakeholders, and identified smartcard technology as a useful and under-researched area. It is the intention that this thesis, when taken as a whole, will provide an improved understanding of MV whilst also demonstrating how new data sources can be applied innovatively to provide evidence on the subject.

### **1.1.1 Policy Context**

In this section it will be shown that this research is timely from a policy perspective. The idea that reliability may have value at least as great as VOT is one that is capturing the attention of practitioners and policy makers, and research activity in this area has increased as a result. The policy focus of this thesis will be the UK, where the research is based and funded, but it is also noted that the UK (along with the likes of the Netherlands) has been at the forefront of relevant policy developments.

The Eddington Transport Study (2006) was instrumental in introducing reliability to the mainstream transport policy debate in the UK by identifying reliability as a key challenge and opportunity for the UK's transport system. It highlighted a requirement to improve the performance of the current transport network by making the recommendation that "*Government action needs to focus on tackling congestion, capacity constraints, and unreliability on existing networks*" (Eddington, 2006, p.58). The summary report makes several points with regards to transportation reliability, that:

- Reliability is valued by users as a key characteristic of the transport system;
- Overall predicted benefits of a project could increase by up to 50 percent in some cases if new evidence concerning the importance of reliability were to be included in the appraisal of transport schemes;
- The importance of reliability is growing, particularly owing to opportunities presented by new technology.

Of particular interest to the present thesis is this final point. Although the Eddington Study is not specific on the nature of the opportunities provided by new technology, the innovative use of smartcard data to improve our understanding of reliability would be consistent with this idea.

Since Eddington in 2006, the UK's Department for Transport has also published '*Supporting economic growth in a low carbon world*' (2007), which reinforced the conclusions of the former regarding reliability: that a reduction in congestion would bring benefits to both business and the environment. Consequently, an emphasis on reliability has become a feature of policy for local and passenger transport authorities and is recognised in many of the UK's local transport plans.

In the Netherlands, the idea of the 'door-to door' trip has become a key part of transport policy (Ministry of Transport, Public Works and Water Management, 2004)<sup>1</sup>. One aim of this policy was to increase the reliability of the entire chain. Accordingly, by 2020, 95% of road based trips will be required to end on-time, and the reliability of the rail network will be similarly enhanced. There was also specific direction provided to local authorities that any change to planning policy that will affect reliability was required be consulted upon with national road and rail authorities.

TfL, a co-sponsor of this thesis, is the authority responsible for the majority of the transport system in the UK's capital city and is accountable to (among others) the elected Mayor of London. In the previous Mayor's Transport Strategy<sup>2</sup>, one of six goals was to "*support economic development and population growth*", under which is the outcome of "*improving public transport reliability*" (Greater London Authority, 2010). It is within this policy context that TfL are seeking to improve their knowledge of reliability on their own transport systems by working with institutions such as the University of Leeds.

## **1.2 Aims and Objectives**

---

<sup>1</sup> The 2004 document is referenced. An English translation of the later "Mobiliteitsaanpak (2008) 'Safely and smoothly from door to door'" could not be accessed

<sup>2</sup> The Transport Strategy of the present Mayor has not been published at the time of writing.

This introductory chapter has suggested that there is some improvement needed to the research on reliability for it to be fully integrated into appraisal in the same way that travel time is at present. The literature review will identify gaps in past research which this thesis will address. Also discussed were the opportunities afforded to this field by the advent of IT-based technology in transportation and the new datasets that these systems generate. The partner organisation to this project, TfL, have agreed to provide smartcard datasets to assist in this study. The aims of the thesis are therefore:

A1. To develop understanding of public transport smartcard data and identify key strengths and limitations through application of the data;

A2. To apply smartcard datasets to the Mean-Variance framework to improve understanding of reliability and passenger responses to it;

A3. To conduct a comprehensive review of the Mean-Variance framework and investigate possible improvements.

These aims will be achieved through the objectives set out in Table 1.1.

**Table 1.1** - Full list of objectives to be met in this thesis

Objective	Meeting Aim:	Description	Addressed in:
O1	A1	To apply smartcard datasets to a number of real world situations, drawing and developing upon existing studies.	<p>Chapter 3 – Datasets</p> <p>Chapter 4 – The factors affecting reliability</p> <p>Chapter 5 – Development of an RP methodology for estimating a VOR</p> <p>Chapter 6 – Utility functions</p>
O2	A1	To provide a critique of the smartcard data available (and smartcard data more broadly)	<p>Chapter 3 - Datasets</p> <p>Chapter 4 – The factors affecting reliability</p> <p>Chapter 5 – Development of an RP methodology for estimating a VOR</p> <p>Chapter 7 - Conclusion</p>



O3	A2	To develop the means for improving understanding of the factors affecting transport reliability using smartcard data	Chapter 3 - Datasets  Chapter 4 – The factors affecting reliability
O4	A2	To develop a methodology for estimating a VOR using smartcard data	Chapter 2 - Literature Review  Chapter 3 - Datasets  Chapter 5 – Development of an RP methodology for estimating a VOR
O5	A3	To review the origins of the Mean-Variance framework from its origins in finance to its transition and use in transport contexts.	Chapter 2 – Literature review
O6	A3	To explore improvements to the standard Mean-Variance framework, including other statistical indicators of risk, the shape of the utility function and potential alternative frameworks.	Chapter 2 - Literature review  Chapter 3 – Datasets  Chapter 6 – Utility functions

It is these aims and objectives that will motivate the work that is to follow in the remaining chapters of this thesis. The relevant objectives will be highlighted at the end of each chapter.

### **1.3 Novelty and Original Contribution**

The primary contribution of this thesis will be toward the literature on reliability; specifically Mean-Variance. This thesis will also make use of smartcard data in new contexts and demonstrate additional capabilities of such data.

The literature review on reliability in Chapter 2 will extend beyond transportation to include research that has dealt with risk in alternative circumstances; namely *portfolio theory* in finance. Although this link has been established and made reference to in the transportation literature (e.g. Carrion and Levinson, 2012), there has been little work in which the transition has been fully explored. This thesis will critique the early work of portfolio theory with respect to its eventual application in transportation. Chapter 2 will also consider the critiques and developments of portfolio theory in finance with a view to cross-fertilising developments to Mean-Variance in transportation. This full consideration of the link between risk in finance and transportation is an original contribution to the literature on the subject.

There will be a significant amount of analysis conducted using the public transport smartcard dataset in this thesis, which represents the study's most substantive contribution to the literature. A key element of this will be to develop a *revealed preference* (RP) method for estimating the VOR using smartcard data. The majority of estimates of the VOR are from SP sources; asking respondents to indicate a choice between two or more scenarios with varying levels of reliability and monetary cost and estimating a discrete choice model based upon the responses. Building on previous reviews of presentational issues related to SP and reliability, an alternative RP approach will be developed in this thesis. Drawing upon a full data description of the smartcard data available that takes place in Chapter 3, the process of applying such datasets for estimating a VOR will be outlined in Chapter 5. This not only represents a contribution to the limited literature on RP estimates of the VOR, but also demonstrates a methodology for similar applications of smartcard data. It represents a unique

application of smartcard data in the context of established choice modelling techniques. Chapter 4 will develop some initial work by other authors that has attempted to estimate regression models to identify the factors driving reliability levels. The corresponding analysis presented in this thesis will incorporate additional data (and explanatory variables) over what has come before. The thesis will also use smartcard data to conduct analysis on alternative risk measures. This is a novel application of smartcard data, and will be further developed in Chapter 6 by using the RP method developed in Chapter 5.

## 1.4 Thesis Outline

Moreover, the thesis is organised as follows:

**Chapter 2** consists of a literature review. This section provides justification for the aims and objectives outlined in Section 1.2 of the present chapter. The review includes sections on expected utility theory, empirical value of reliability studies and provides a critique of the research both in terms of methods and outcomes. A key element of the chapter is to compare SP and RP methods and thereby develop the argument for undertaking further RP analysis in Chapter 5.

The first section of **Chapter 3** is primarily descriptive in nature and gives a fuller account of the datasets available to the study. This is focussed on public transport smartcard datasets. The second section is based upon initial analysis and identifies potential issues with using these data. The third section builds upon the initial analysis by using the smartcard data to compare candidate MV reliability indicators and identify those suitable for further analysis in Chapter 6.

**Chapter 4** makes use of the key smartcard datasets to calculate a range of explanatory variables that may be related to reliability levels. These are used to estimate a linear regression model explaining changes in reliability levels. Another use of these models would be to predict the standard deviation of travel time. The extent to which the models developed are capable of this is reflected upon.

**Chapter 5** exploits the smartcard datasets as an RP data source. It is shown how MV parameters can be estimated based on these data. Two choice model specifications

are introduced and estimated using the data. The model results are presented and discussed both in relation to the VORs obtained and the data used.

**Chapter 6** builds upon the analysis conducted in Chapter 3 to identify alternative indicators of travel time risk. It does so by applying the methodology developed in Chapter 5 to each of the indicators in turn. A discussion also takes place regarding the shape of the utility function implied by alternative indicators to the standard deviation. This leads to the final element of the study: a comparison of the shape of the MV and Scheduling utility functions and whether they imply the same behaviour on the part of the traveller.

In **Chapter 7** an overview of the work undertaken is given and the potential impacts of the thesis upon the fields of smartcard research and transportation reliability are evaluated. The thesis is further evaluated against its limitations, as well as the stated aims and objectives of the research, and possibilities for future research are discussed.

## **Chapter 2 - Literature Review**

### **2.0 Introduction**

This literature review chapter builds upon the introduction provided in Chapter 1. In Chapter 1, reliability was introduced as an active research area, one where unresolved questions remain. Chapter 1 outlined the aims and objectives for this thesis in the light of a general discussion of the field. In the present chapter, these aims and objectives will be justified through a detailed review of the research that has already been conducted. This chapter will also provide an initiation to the main underpinnings and active areas of research which will be referred to in the subsequent five chapters of the thesis. Although the smartcard data that was introduced in Chapter 1 will form the basis of the analysis that is to come, the literature review is focussed on estimating the value of reliability. Particular emphasis is placed upon research that focuses on public transport where available.

The first substantive section of this chapter is introductory and will concentrate on the microeconomic framework underpinning much of the reliability research; expected utility theory (EUT). The origins of this area will be outlined, and the description will then move onto its application in transportation. The background related to risk attitudes of travellers will be of particular interest for the work that is to follow. Some research has recognised issues with EUT, and these will be identified and commented upon. However, it is EUT that will be the framework that underpins the substantive analysis in later chapters.

Against this background, the Mean-Variance (MV) approach will be introduced in detail. MV is the focus of the work that follows and therefore this section will form a key part of the present chapter. This section will be chronological in nature by tracing MV back to its origins in the field of finance and portfolio theory (Markowitz, 1952; 1959). Some additional research by those in the field of finance will also be introduced, before MV's transition into transportation contexts (by Jackson and Jucker, 1982) is outlined and critiqued. This section will conclude with an investigation into further developments to the theory of MV made by researchers in the field of transportation

studies. This will particularly focus upon alternative statistical indicators to represent reliability which is a theme developed in later chapters.

Prior to a review of empirical evidence on the valuation of reliability, alternative approaches to MV are outlined. These include the aforementioned Mean-Lateness and Scheduling approaches. The Mean-Lateness approach is of interest as it is well developed and utilised within the UK rail industry. The Scheduling approach too has been developed within transportation research. The development of these frameworks will be briefly narrated and critiqued. The final part of the frameworks section will explore the theoretical work which has established equivalence between MV and the Scheduling approaches

With the microeconomic underpinning of transportation reliability outlined, and the three primary frameworks for its valuation introduced, the chapter will then provide an overview of the empirical evidence on the value of reliability (VOR) that has been generated through application of these frameworks. The initial focus of this discussion will be Stated Preference (SP) survey methodologies. This will be progressed through review of SP evidence and discussion of the issues related to this approach. The alternative method of data collection to value reliability is Revealed Preference (RP). Again, review of methodological issues will form the first part of the RP section followed by valuation evidence. The valuations obtained from both SP and RP will be summarised and compared, reporting these as Reliability Ratios where possible.

The final substantive element of the literature review will return to the subject of equivalence between MV and Scheduling frameworks and examine the research that has attempted to establish this link in real-world situations.

Based upon the above, the remainder of this chapter therefore takes the following form: Expected Utility Theory is outlined in Section 2.1; the introduction, contextualisation and critiques of MV are provided in Section 2.2. In Section 2.3 the alternative Mean-Lateness and Scheduling frameworks for the valuation of reliability are formally introduced as well as the idea that MV is theoretically equivalent to the Scheduling framework. Section 2.4 focusses on the empirical issues and outcomes relating to these frameworks; initially focussing on SP and progressing on to RP, with a

comparison of the evidence obtained using each method. The section ends with a review of empirical evidence related to equivalence between MV and Scheduling. Arising from the literature review, the original aims and objectives will be contextualised in the concluding section of the chapter.

## 2.1 Expected Utility Theory

Expected utility theory (EUT) is the microeconomic framework upon which most of the transport reliability literature is based and will be the underpinning of the research conducted in this thesis. Broadly speaking, it is the theory that a traveller will maximise their expected utility when faced with travel time risk. It will be useful to unpack this statement in greater detail.

The term ‘risk’ can be defined as a situation where the full range of possible travel times is known to the traveller, and furthermore the probability of each of these travel times occurring is also known. This is in contrast to the term ‘uncertainty’ where the probability is unknown to the agent (Knight, 1921). In the case of a discrete distribution of travel times, the traveller will under EUT attempt to maximise the function:

$$EU = \sum_{k=1}^K P(t_k) \cdot U(t_k) \quad (2.1)$$

Where  $EU$  is the expected utility,  $t$  is the travel time and  $k$  is an indexation of each of the (discrete) possible travel times so that  $P(t_k)$  is the probability of travel time  $k$  occurring.  $U(t_k)$  is the utility resulting from  $t_k$ . The summation over all  $k$  in Equation 2.1 therefore results in the expected utility. It could be argued that  $t_k$ , representing time, would be a continuous variable, therefore requiring an integration of its associated probability distribution; however its treatment as discrete is sufficient for the work that follows.

It is important to understand the intuition behind the concept of expected utility,  $EU$ . This can be explained with reference to the St Petersburg paradox and the work of Bernoulli (1954<sup>3</sup>). The paradox is based upon a game where a fair coin toss takes place and a prize of two pounds is received by the player if ‘heads’ results. If ‘tails’ results,

---

<sup>3</sup> A reprint of the work which was originally published in 1738

the coin is flipped again and the prize is doubled to four pounds if ‘heads’ is the result. The prize is doubled every time ‘tails’ is flipped but ends when ‘heads’ results. If the number of coin tosses is denoted by  $n$ , the expected value of the game is:

$$E(\pounds) = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8 + \dots + \frac{1}{2^n} \cdot 2^n \quad (2.2)$$

As  $n$  can in theory be infinitely large, the expected outcome is infinite and individuals should be willing to pay a large sum of money to take part in this game, yet this is not the case in reality. Bernoulli suggested that individuals do not maximise the expected value of the risky choice, but rather its expected utility. It was further proposed that the relationship between expected payoff and utility would be non-linear, reflecting a diminishing marginal utility of wealth and risk aversion. Arrow (1974) succinctly defines a risk averse individual as:

“one who, starting from a position of certainty, is unwilling to take a bet which is actuarially fair” (Arrow, 1974, p.90).

Von Neumann and Morgenstern (vNM) (1947) formalised EUT in the development of game theory, the key result of which was four axioms of preference that are necessary for EU maximising behaviour. The first of these axioms is that of *completeness* where an individual has well defined preferences and is able to choose between outcomes. This condition therefore implies ‘risk’ as opposed to ‘uncertainty’ – risk being the circumstance where the probability of each of the outcomes is known in advance as opposed to uncertainty where these are not known. If these outcomes are defined as  $t_1$  and  $t_2$ , representing two travel times for example, then:

$$t_1 \succcurlyeq t_2 \text{ or } t_2 \succcurlyeq t_1 \quad (2.3)$$

Where  $\succcurlyeq$  means ‘is at least preferred as’. The two conditions in Equation 2.3 do not preclude the two travel times being equally preferred. The second vNM condition is *transitivity*, which is essentially consistency of preference ordering over prospects:

$$\text{If } t_1 \succcurlyeq t_2 \text{ and } t_2 \succcurlyeq t_3 \text{ then it is implied that } t_1 \succcurlyeq t_3 \quad (2.4)$$



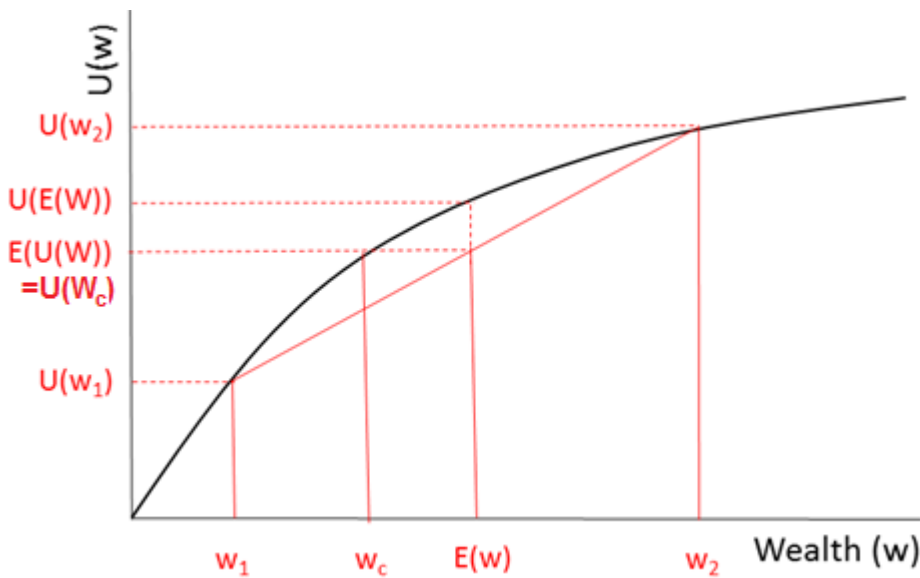
The third axiom of EUT as proposed by vNM is *continuity*, which is probabilistic in nature. If  $t_1 \succeq t_2 \succeq t_3$ , then there exists some probabilistic combination of  $t_1$  and  $t_3$  such that an agent would be indifferent between that combination and  $t_2$  alone.

The fourth axiom is that of *independence*; where the preference ordering between two alternatives is maintained when each is separately combined with a third independent alternative with the same probability value,  $p$ . That is:

$$\text{If } t_1 \succeq t_2, \text{ then } p(t_1) + (1 - p)(t_3) \succeq p(t_2) + (1 - p)(t_3) \quad (2.5)$$

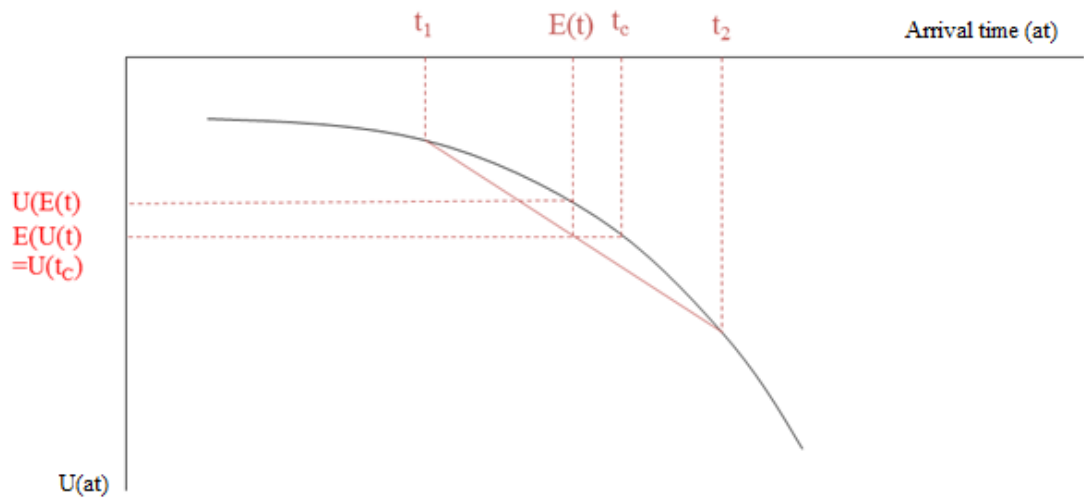
Further detail on these axioms beyond vNM (1947) can be found in Savage (1954) and Fishburn (1970).

EUT also yields insights concerning an agent's attitude towards risk. Bernoulli (1954) recognised that the marginal utility of wealth decreases as wealth increases: that an extra unit of wealth is less valuable to a person with high wealth as compared with low wealth. Therefore Bernoulli suggested that the relationship between wealth and utility of wealth was concave. A concave utility function was shown separately by Arrow (1970) and Pratt (1964) to be indicative of aversion to risk. This is demonstrated in Figure 2.1 for a risky prospect of two wealth levels,  $w_1$  and  $w_2$ . This situation shown is analogous to a fair coin toss with 'heads' resulting in  $w_1$  and a 'tails'  $w_2$ , so that  $p(w_1) = p(w_2)$ .



**Figure 2.1** - Illustration of a concave (risk averse) utility function and featuring a single risky prospect with two outcomes of equal probability

Figure 2.1 shows a concave relationship between wealth and the utility of wealth. A prospect is defined as a gamble where each outcome of wealth is associated with a probability ( $w_1, p_1; \dots; w_k, p_k$ ) and  $\sum p_k = 1$ . Therefore in the example above, only two values of wealth are shown, given that there are only two outcomes. If the player loses the coin toss their resulting wealth is  $w_1$ ; a win takes their wealth to the point  $w_2$ . As the coin is fair, the expected wealth  $E(w)$  resulting from the bet is the halfway point  $w_1$  and  $w_2$ . The curvature of the utility function results in the expected utility of  $W$  being lower than the utility of the expectation of  $W$ . Of interest to this thesis (particularly Chapter 6) is the value  $w_c$ , the so-called ‘certainty equivalent’, which is at a lower level of wealth than  $E(w)$ . The difference given by  $E(w) - w_c$  is referred to as the ‘risk premium’ and represents the amount the player might be willing to pay to avoid the situation involving risk. The concave utility function has been translated into transportation contexts so that wealth is replaced by travel time (Polak, 1987a). This is shown in Figure 2.2.



**Figure 2.2** - Illustration of a concave (risk averse) utility function and two risky travel time outcomes with equal probability

Figure 2.2 is in the negative domain of utility as travel time is most often seen as a ‘bad’ by travellers. Only two travel time outcomes are shown,  $t_1$  and  $t_2$ , where  $t_1$  is a short travel time and  $t_2$  is longer. In this example the probability of each occurring is assumed to be 0.5 (analogous to the coin toss), and furthermore the departure time is assumed to be fixed. It should be noted that in travel time contexts there will often be more than two outcomes.

In Figure 2.2 the utility of the expectation of  $t$  remains greater than the expected utility of  $t$ . Unlike in the example of wealth, the value of the certainty equivalent  $t_c$  is in excess of  $E(t)$  on the x axis. This difference between  $t_c$  and  $E(t)$  is the willingness of a traveller to pay in units of travel time,  $t$ , to avoid risk in travel time (Batley, 2007). In both examples, if the utility function were linear this would imply a neutral attitude towards risk, and convex curvature would imply risk seeking behaviour where the agent would be willing to pay something (in money or time) in order to participate in the gamble.

Despite the apparent flexibility of this approach, it will be worthwhile to note some negative aspects of EUT that have been raised in the literature. Deaton and Muellbauer (1980) questioned whether an approach based upon expected values and expected utility is reasonable given that in many cases the risky situation will be only encountered once by the agent, and therefore expected values are not of utmost importance to them –

rather it is the actual outcome on a given occasion. A second concern raised by these authors is more fundamental: whether probabilities can be attached to many events in reality? To illustrate this point the authors contrasted a game of chance, where probabilities can be defined, to a question such as whether there will be a man on the planet Mars by a given year, where the probability associated with this question is unknowable. A further question is to what extent individuals are able to correctly assign probabilities to events? The example given by Haber and Runyon (1973) is of 30 individuals in a room who are asked to bet on the event that two or more have the same birthday. The actual probability of this event is 70%, although it is unlikely that this result would be readily accepted by the participants.

There are also concerns raised in the literature about the nature of the utility function, particularly the common assumption of a form similar to that shown in Figure 2.1. Friedmann and Savage (1948) suggested that the utility function would not be completely convex or concave, with agents exhibiting different attitudes to risk at different levels of wealth. This is an idea that will be returned to in Chapter 6 of this thesis. That agents might have multiple attitudes toward risk is a key part of 'Prospect Theory' (Kahneman and Tversky, 1979). This prominent work was based upon finding that the axioms of EUT are often violated in reality (Allais, 1953).

One criticism of this example in Figure 2.2 might be that two arrival time outcomes are unlikely in the context of travel time as time is a continuous variable. Batley (2007) justifies the use of a two outcome prospect in the context of travel times:

- von Neumann and Morgenstern's axioms of expected utility refer entirely to the properties of two outcomes
- Two choice options are commonly utilised in SP choice studies within the transportation field

Many have concluded that EUT is a sufficient framework upon which to base further research, particularly in the field of transportation reliability (Bates et al, 2001). It is EUT that underpins MV, the primary framework for the valuation of reliability in transport contexts and the focus of the next section.

## **2.2 Mean-Variance: Background and Development**

Mean-Variance (MV) represents the primary approach in transportation for the valuation of reliability. However MV traces its foundation to the field of finance under the heading of *portfolio theory*. Portfolio theory has proved to be attractive to the field of transport in that it deals with choices made under risk. Originally formulated by Markowitz, with supporting contributions from eminent economists such as Marschak, Keynes and Tobin, the MV approach is well established in finance. Broadly speaking, its treatment of risk is analogous to reliability in transportation applications. Researchers in the field of transport have tended to attribute this link to Jackson and Jucker's (1982) empirical exposition of MV in a transportation context.

This section provides an initiation into MV for researchers in the field of transport. It will begin by providing an outline of the early development of Portfolio Theory. This will then be contrasted with the motivations and applications of the MV approach to reliability in the field of transportation – notably the work of Jackson and Jucker (1982) and Senna (1994). The reader will be made aware of inconsistencies between the two fields, which opens the way for an examination of possible alternative measures of risk in Chapters 3 and 6. The section is drawn to a close with a concluding discussion of the applicability of MV to transportation reliability in advance of the empirical work to come.

### **2.2.1 Portfolio Theory: Early Development**

To understand the early development of portfolio theory it is useful to discuss the key questions behind the work of Markowitz and others in the field of finance. In understanding this motivation, it is hoped that the applicability of MV to transport reliability will become more apparent.

The reader versed in the transport literature would be forgiven for thinking that risk was central to the development of MV in finance: with MV providing an answer to the question “how do investors respond to risk?”. In fact, Markowitz (1952, 1959) and others were more interested in explaining how investors allocated wealth within financial markets, in particular choosing between the holding of wealth as cash (assumed to be a riskless asset with no return), as an investment portfolio comprising a

number of securities, or a combination of the two. The problem can be formalised in an example of a single investor with initial wealth  $A_0$  and a set of  $L$  (discrete) investment opportunities denoted by  $\{l_1, l_2 \dots l_N\}$ .  $A_0$  is apportioned to the members of  $L$  such that  $a_1$  is the proportion of wealth invested in investment opportunity  $l_1$  and  $\sum_i^N a_i = 1$ . The expected payoff of each member of  $L$  is denoted by the set  $W = \{w_1, w_2, \dots, w_N\}$ . Assuming an ordering of such payoffs  $w_1 > w_2 > \dots > w_N$  and an investor focussed on only maximising the expected payoff, then  $a_1 = 1$  i.e. the initial wealth is invested entirely in the single fund or bond that has the highest payoff. However it was known from observation that investors did not invest their entire wealth in a single, high payoff fund; and consequently some other phenomena needed to be taken into account. Markowitz recognised that investors would maximise the expected outcome whilst minimising the risk involved in the portfolio. This is the key contribution that was later recognised in transport: there is a trade-off between the expected outcome and the risk associated with that outcome.

Prior to the breakthrough of Markowitz (1952), research in this area was focussed on what is termed “liquidity preference”, notable examples include the work of Keynes (1936) exploring the phenomenon of agents holding their wealth as cash as opposed to investments; despite the inability of cash to generate returns on this capital. A feature of cash assets in this context was that it was assumed to be “riskless”, whereas investments in securities carried with them an element of risk. The early developments of the “liquidity preference” therefore sought to explain the allocation of wealth between risky and non-risky assets.

Early work by Hicks (1935) recommended that a measure of dispersion was necessary for characterising the risk associated with investment, although this was not formalised until later. In his work it was suggested that agents would trade off between a preference for an increased expected payoff and decreased standard deviation of that payoff.

Marschak (1938), in advance of the framework provided by von Neumann and Morgenstern (1947), developed the concept that agents have a probabilistic approach to risk – that future events had a range of outcomes and attached to each of these was a

known probability. To characterise this formulation of risk, a number of variables were introduced in the work, including the expected yield of an investment along with its standard deviation. Marschak also mentioned correlations between alternative investments as important, which is a key component of Markowitz's Portfolio Theory, although did not develop further upon this idea.

The correlation coefficient was crucial to the first major work of Markowitz (1952), who provided an exposition of MV as a means to explain observed behaviours in the selection of portfolios. The work initially addressed two issues; the first being the idea that investors seek to maximise returns only, and the logical outcome of this being their placing of all wealth in a single fund with the highest payoff. The second issue was crucial to the contribution of the 1952 work – that is, why investors will not simply spread their capital among many securities and rely on the law of large numbers to effectively hedge against risk. Ignoring transactional costs (as Markowitz did at this point), the answer to this question is covariance between securities – as is intuitively put in the 1952 work:

“A portfolio with sixty different railway securities ... would not be as well diversified as the same size portfolio with some railroad, some public utility, mining, ...” (Markowitz, 1952, p.89).

Markowitz found that investors could maximise the expected utility of their portfolio by taking into account the expected outcome and risk only. The expected outcome of the portfolio, defined as  $E$ , was given as:

$$E = \sum_{i=1}^n a_i \mu_i \quad (2.6)$$

Where  $\mu_i$  is the mean return of security  $i$ . The variance of the return of such a portfolio was given by:

$$V = \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} a_i a_j \quad (2.7)$$

Where  $\sigma_{ij}$  is the covariance between securities  $i$  and  $j$ . In both cases, the maximisation would be subject to the allocation vector,  $a$ .

This solution does not however imply the same portfolio would always be chosen given the same range of securities – this is determined by the preference parameters of  $\eta$  and  $\rho$ . Assuming a fixed trade-off, the investor would therefore maximise the following expected utility function:

$$EU = \eta E + \rho V \quad (2.8)$$

Where  $E$  is the mean return and  $V$  the variance of return. As agents would wish to maximise  $EU$ , Equation 2.8 illustrates that this will be a process of trading between the main return and risk associated with that return. Of relevance to the later discussion of MV and transportation, the variance is commonly replaced by the standard deviation (i.e. the square root of the variance). Markowitz was unconcerned by this change:

“Of course, any relationship expressed in terms of variances can be translated into terms of standard deviation by substituting (standard deviation)<sup>2</sup> for variance.” (Markowitz, 1959, p.82)

In a later chapter of the same work, he outlines justification for use of the standard deviation as a measure of risk:

“... if an investor

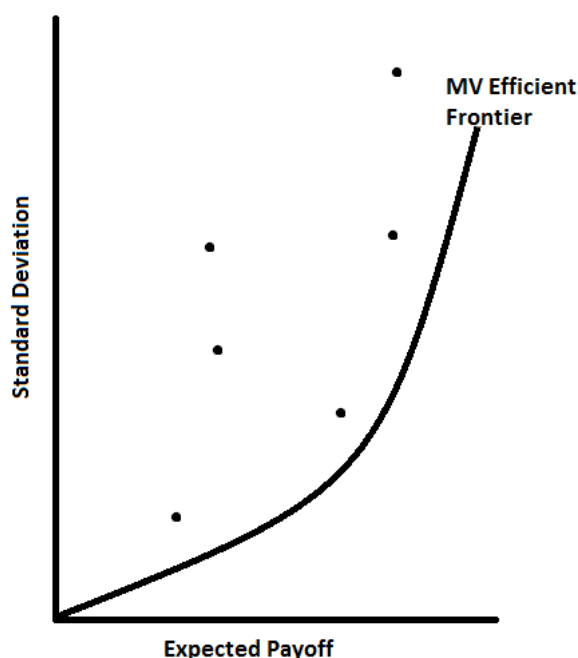
- a) Maximises the expected value of some utility function, and
- b) His choice among portfolios depends only on their expected returns and standard deviations

Then his utility function must be quadratic.” (Markowitz, 1959, p.288).

Markowitz developed this argument to say that if the quadratic was sufficiently close to the investor’s true utility function, then a mean-standard deviation efficient portfolio would maximise their expected utility. This support for use of the standard deviation goes some way to justifying its common use in investment and transportation contexts.



Returning to Equation 2.8, Markowitz quite reasonably assumed that investment firms saw return as a positive and risk as a negative – i.e.  $\eta > 0, \rho < 0$ . Markowitz’s earlier paper on portfolio selection also introduced the concept of an efficient frontier of portfolios. As previously described, the investor was able to allocate starting capital  $A_0$  across available securities. As  $a_N$  was defined as a continuous variable, it should be clear that a large number of combinations of securities were possible, not all of which would be efficient – i.e. maximising expected payoff for a fixed level of risk or minimising risk for a fixed payoff. However, it was possible that a number of portfolios of securities may be equally efficient, in which case an “efficient frontier” was defined where portfolios were at least as good as one another. Portfolios on this frontier could be chosen according to the agent’s preferences related to payoff and risk. This frontier was presented graphically by Markowitz (1959) and is reproduced below (Figure 2.3). Each point represents a single portfolio. For any portfolio on the efficient frontier an increase in expected payoff results in an increase in the risk of the portfolio (here denoted by the standard deviation). Similarly if an agent wishes to decrease the risk associated with their portfolio then they must accept a reduced expected payoff.



**Figure 2.3** - Expected payoff and standard deviation combinations with an efficient frontier line (reproduced from Markowitz, 1959)

In this section the motivating question behind the establishment of portfolio theory has been outlined and the basic application in a finance context has been shown.

### **2.2.2 Mean-Variance and Transportation Reliability**

In this section the link between MV in finance and transportation will be examined. The section will not seek to explicitly accept or reject the use of MV in transportation, but rather will provide commentary on the link between the two fields. The purpose of the work is to clarify the roots of MV in transportation – whether transportation researchers and practitioners can look back to 1952 and Markowitz, or should instead focus on more recent work in the field of transportation itself. Such a motivating question may appear trivial and one of historic interest, however the lack of theoretical work on the foundation of MV in the transportation field might imply a reliance on the eminent work of Markowitz, Tobin and Marschak to justify its existence.

Firstly, it will be useful to answer “why has the use of MV come to be so readily accepted in transportation?” The first possible reason is an intuitive acceptance of the standard deviation as a proxy for risk. Markowitz (1959) deliberated on a number of measures of dispersion to fulfil this role, but in transportation it is commonly the standard deviation which is used. Secondly, it could be speculated that the theoretical foundations laid down by portfolio theory, drawing upon the celebrated work of von Neumann and Morgenstern, Markowitz, Tobin and others, have paved the way for ready acceptance of MV.

Leaving such contentions aside, it makes sense to return to the motivation of the finance literature in developing portfolio theory. To recap, the motivation is not to explain risk-return on individual investments, but rather to explain diversification of investments and how to conduct this process efficiently given the existence of covariance between individual investments. As currently applied, there is no analogous motivation for MV in transportation (perhaps with the exception of work by Levinson and Zhu (2013)). It is clear that it is not straightforward for travellers to spread risk among a number of alternatives. If the case of a traveller making a multi-modal trip is considered, then it is difficult to imagine that this is done for reasons of spreading risk between modes – indeed travellers would often seek to minimise interchanges as they

seek to minimise risk on their trip (Noland and Polak, 2002). Another analogy of diversification in transport might be a car user employing a number of proximate arterial roads on their commute in order to ‘diversify’ their route and spread risk – their route in this context is the portfolio, with the constituent links analogous to securities. Such a comparison has not been investigated in the transport context however. In other words, it might be argued that the ‘portfolio’ part of portfolio theory is lacking from mainstream analysis in the transportation context.

In the transport reliability literature, an explicit link has been made between portfolio theory in finance and MV in transport. Examples of this can be found in Senna (1994), Batley and Ibáñez (2012) and Carrion and Levinson (2012). However, many of the other prominent works in this field make no reference to the origins of MV in finance. Although it would seem that there is no consensus in the literature on this point, most authors in the field appear to cite the work of Jackson and Jucker (1982) as responsible for the initiation of MV into a transportation context. Jackson and Jucker themselves cited two resources in the course of introducing MV to transport: a working paper of Brastow and Jucker (1977), and a report by Abkowitz et al (1978). The former has attracted little attention due to lack of publication and consequent difficulty of access. The latter work of Abkowitz et al provided the first formal consideration of variance; focussing on public transport operations from both the point of view of operator and passenger. It should be noted that this report did not introduce MV *per se*, but rather considered all possible options for quantifying TTV. The recommendation of the report was that mean and standard deviation (through the coefficient of variation) should form the basis for measuring reliability. The standard deviation has been the most common measure of reliability in transportation, but this is unsurprising given Markowitz’s (1959) recommendation that it is:

“easier to use, more familiar to many and perhaps easier to interpret” (Markowitz, 1959, pp77).

Markowitz also suggested that other measures may produce superior portfolios, and merely described the standard deviation as “satisfactory”. Nevertheless the decision to use the standard deviation as a measure of reliability in transport is not final and

alternatives have been investigated (e.g. van Lint et al, 2008). This area will be further expanded upon in Chapters three and six.

Abkowitz et al (1978) also recognised the skewed nature of the travel time distribution and therefore recommended a third measure (to add to the mean and standard deviation) to identify the proportion of observations above 2.32 standard deviations from the mean. This acknowledged the impact of extreme lateness upon travellers. In accounting for three moments of the travel time distribution, this early proposal could be said to be an advance upon the two moment approach which has come to be the norm. However, it should be noted that the work of Abkowitz et al (1978) was an applied report written for practitioners involved in the provision of public transport. The work did not make reference to an underlying microeconomic framework but rather proposed a statistical method for monitoring reliability. It also remains unpublished in an academic context. It is for these reasons that credit for the introduction of MV in transportation has generally not been assigned to Abkowitz et al (1978), but rather to Jackson and Jucker.

Examining the work by Jackson and Jucker (1982) in more detail, the model proposed for each traveller,  $i$ , is defined as:

$$\text{minimise } E(T_p) + \lambda_i V(T_p) \quad (2.9)$$

With

$E(T_p)$  as the expected travel time on path (route)  $p$

$V(T_p)$  as the variance of the travel time on  $p$

$\lambda_i$  as a risk aversion coefficient ( $\lambda_i > 0$ )

For  $p \in P$

Implied in Equation 2.9 is the assumption that expected travel time and variance of travel time are fixed for each route  $p$ , and the common assumption is made that it is known by all travellers (i.e. a risky situation as opposed to an uncertain one). Of most interest is the variance parameter  $\lambda$  which is free to vary between travellers. The model

therefore recognises the trade-off passengers would contemplate between mean and variance of travel time. The research used data from an SP questionnaire to calculate a sample distribution of  $\lambda$ , and found evidence of risk aversion. The authors concluded that their method of model calibration was reasonably reliable for predicting the trade-off between time and reliability. They noted the difficulty of presenting the concept of reliability to travellers who in general do not think in purely statistical terms; this remains a live topic within the transport literature to the present day (and will be considered further in Section 2.4).

Many studies have since referred to work of Jackson and Jucker (1982) in order to develop their own methodological or empirical approaches to reliability in the transport context. Key developments of MV from this initial work were made by Polak (1987a, 1987b) and Senna (1994).

Polak (1987a, 1987b) was the first to explore the theoretical side of MV in a transportation context; recognising the translation of the utility function from return on investment to travel time. Polak (1987a) therefore provided the foundation for much of the subsequent work in transportation. A further contribution of Polak was an attempt to bring the theoretical underpinning of MV in transportation up to date *vis-à-vis* the finance literature. This was primarily achieved through the consideration of polynomial and exponential forms of the utility function as an alternative to the quadratic (Polak, 1987b; similar to the work of Tobin, 1958, among others). Nevertheless it is fair to say that these alternative specifications to the quadratic have largely remained unused. Senna (1994) recognised the restrictions of the MV approach already identified by the finance literature (Richter, 1960); specifically, for a utility maximising traveller one of the following must be true:

1. The utility function must be quadratic
2. The distribution of travel time must be Normal

The proof that the variance and a quadratic shaped utility function are related was given in Markowitz (1959) and is reproduced here:

The variance of a distribution can be expressed as a function of  $t$  and  $E(t)$ :

$$\sigma^2(t) = E[(t - E(t))^2] \quad (2.10)$$

Which expanded gives:

$$\sigma^2(t) = E(t^2) - 2 \cdot E[t \cdot E(t)] + E[E(t)]^2 \quad (2.11)$$

And simplifies to:

$$\sigma^2(t) = E(t^2) - [E(t)]^2 \quad (2.12)$$

It is now possible to substitute this equation for variance into the general quadratic form ( $ax^2 + bx + c$ ) to give:

$$EU = a [\sigma^2(t) + [E(t)]^2] + bE(t) + c \quad (2.13)$$

The second restriction, of MV to a Normal distribution of returns on investment, was recognised as problematic in the finance literature because returns on investments were rarely Normally distributed (Samuelson, 1970; Tsiang, 1972). Nevertheless MV was largely accepted as a ‘sufficiently useful’ framework for the efficient diversification of investments. In transportation it is unusual to observe a Normal distribution of travel times; more common is a positive skew (Casello et al, 2010; Fosgerau and Karlström, 2010; Rietveld et al, 2001).

Senna (1994) raised the possibility of a risk prone traveller by suggesting that  $\lambda$  could take non-positive values – something explicitly rejected by Jackson and Jucker (1982). Senna’s key contribution was a more formal treatment of MV, where mean and standard deviation of travel time were treated within a microeconomic framework. This theoretical discussion of risk was demonstrated with standard two-outcome prospects to establish the existence of a certainty equivalent for risk-averse, risk-neutral and risk-seeking travellers. Senna also drew upon earlier research in both transport and finance to briefly discuss the functional form of a traveller’s underlying utility function. Some discussion of the superficially attractive quadratic utility function was provided, primarily based upon the finance literature of the 1950s and 1960s. Had Senna’s literature review of this subject extended further, his work might have noted the

rejection of quadratic utility by some researchers within the finance literature (e.g. Tsiang, 1972). Senna also introduced the possibility of additional moments of the travel time distribution being more suited to a polynomial utility function. This insight supported the intuition behind the three parameter model of Abkowitz et al (1978).

Beyond Senna (1994), development of the MV model in transportation has been limited. Much of the work since has instead focussed on empirical application of MV or its interaction with the Scheduling approach. One key development in relation to the applied literature has been the emergence of the *Reliability Ratio* (RR) (Black and Towriss, 1993). This was defined as the ratio of the marginal utility of travel time risk (in this case represented by the standard deviation of travel time) to the marginal utility of travel time. It can be defined as follows:

$$RR = \rho/\eta \quad (2.14)$$

Where  $\rho$  is the marginal utility of standard deviation, assumed to be negative, and  $\eta$  is the marginal utility of travel time, also negative.

Example values of the RR will be provided in Section 2.4 of the present chapter, and further estimation of the RR will be the focus of the empirical work that takes place in Chapter 5.

### 2.3 Mean-Variance Indicators

Markowitz (1959) expressed concern that the mean and variance were not the most appropriate measures to use in his framework. In this section the idea of alternative indicators, particularly for risk, will be introduced. Alternative statistics for travel time risk will be investigated using smartcard data in Chapter 3.

Markowitz initially addressed this point by considering the common alternative measures of central tendency to the mean; being the *mode* and *median*, in relation to both Normal and Lognormal distributions. Markowitz concluded that the mode was too sensitive a measure, whilst the opposite was true of the median, and although mean was carried through in the analysis, other measures might be more appropriate in some instances.

Markowitz also devoted significant attention to alternatives to the common measure of risk used; for example Chapter 9 of Markowitz (1959) was devoted to semi-variance (SV). Semi-variance in that context was described as the variance of values that fall below a certain target of investment return. In the context of travel time reliability it can be defined as:

$$SV = \frac{1}{n} \sum_{t_i > \mu}^n (\mu - t_i)^2 \quad (2.15)$$

Where  $t_i$  is the travel time for traveller  $I$ ;

$\mu$  is the mean travel time; and,

$t_i > \mu$  limits the calculation to cases where travel times are greater than the mean.

Chapter 13 of Markowitz (1959) extended this work with four other possible measures of risk such as “expected value of loss” and the “maximum loss”, thereby continuing the theme of focussing on losses rather than gains as the main motivation of agents. In doing this, Markowitz was looking for the most appropriate measure of dispersion under the assumption of a concave utility function. Most measures of spread were demonstrated to result in a linear utility function in investment losses – suggesting a neutral risk attitude in losses which is intuitively unreasonable and contrary to Markowitz’s original assumptions. Markowitz also briefly considered the quadratic that reaches a maximum within the range of expected returns, thereby suggesting some unfeasibly large aversion to high payoffs beyond the maximum point.

Markowitz showed that the semi-variance possessed appealing characteristics as a measure of risk. In that discussion, it was demonstrated that the implied utility function was linear in gains, and concave in losses (i.e. which would deviate from the properties of Figures 2.1 and 2.2 above). It was asserted by Markowitz that agents may be less risk averse in gains than losses and therefore linearity (or perhaps even convexity) in gains was not prohibitive. This proposition seemingly allows for a utility function that exhibits multiple risk attitudes, such as those proposed by Friedman and Savage (1948). Despite this insight, it is notable that the semi-variance is rarely used in MV in any context, whether finance or transportation.



What remains unclear from Markowitz (1959) is how measures of centrality and dispersion should be selected. The work of Jackson and Jucker (1982) considered a standard range of measures similar to Markowitz. They followed Markowitz in assuming the agent's utility function was approximately quadratic and therefore their expected utility was a function of mean and variance of travel time. They also suggested that the use of variance would assist in the transport context due to its ability to be summed over road or public transport links to create a single value to represent risk on an entire route. What was clear from both the founding work in finance and transportation is that the measures of centrality and dispersion are not necessarily fixed.

This debate on alternative indicators of dispersion is one that remains unresolved in the transportation context. Lam and Small (2001) and Small et al (2005) favour a percentile based indicator of reliability, as this has the ability to account for a skewed distribution, whilst also omitting extreme travel time observations that might be attributed to atypical travel behaviour. On the other hand, Hollander (2006) found that such a percentile indicator offered a poorer account of choice behaviour in his study. A literature has emerged which attempts to identify the most appropriate indicator of travel reliability (Bogers et al, 2007; Bogers et al, 2008; van Lint et al, 2008). Although such research has recognised the importance of skew in explaining traveller behaviour, no clear replacement to the standard deviation has emerged.

#### **2.2.4 Conclusion**

Section 2.2 has provided an outline of the key developments in the theory of MV from its origins in finance as a core aspect of portfolio theory, to present usage in the field of transportation. The crucial point of this development was Jackson and Jucker (1982), which was one of the first applications of MV in a transportation context. Section 2.2.1 was exclusively focussed upon MV in finance contexts, where a description of its original purpose identified the key element of MV relevant to the transportation context – namely the trade-off between expected return and risk. Section 2.2.2 was focussed upon the work of Jackson and Jucker (1982) and how well the transition from finance to MV is explained by this work. It was shown that the re-establishment of the link between MV in transportation and EUT can be attributed to Polak (1987a, 1987b) and Senna (1994). Section 2.2.3 contrasted MV in portfolio

theory and transportation, and raised questions on aspects of portfolio that have not been translated to transportation context, such as choice of statistical indicators.

Unlike MV, the Scheduling and Mean-Lateness approaches to reliability originated in a transportation context. A commentary on the development of these approaches is provided in the next section.

## **2.3 Alternative Reliability Frameworks**

### **2.3.1 The Scheduling Approach**

The Scheduling approach is the primary alternative to MV in the literature. A key feature of the approach is the assumption of a *preferred arrival time* (PAT) on the part of the traveller, such that arrival at the destination before or after this ideal arrival time incurs some disutility. The necessity of a PAT may reduce the generality of the approach in relation to MV, as there may be types of travel where a PAT is less well defined; for example, leisure trips. There have been attempts to account for flexibility in travellers' schedules (e.g. Janelius, 2011), nevertheless the literature has tended to focus upon commuting trips with a fixed PAT. Such a case represents the situation where the commuter expected to begin work at a specific time. The Scheduling approach also differs from MV insofar as it explicitly treats departure time as the key method for travellers to deal with uncertainty in their travel times: a traveller chooses a departure time which maximises their EU subject to their dislike of arriving early or late.

The first major contribution to the founding of the Scheduling approach was Gaver (1968). Gaver assumed a specific point in time at which the traveller would like to arrive at their destination, i.e. the PAT. Crucially, the work also recognised a different cost was associated with early and late arrival at a destination. This formally introduced a degree of asymmetry into the framework around a reference point (the PAT). Initially the assumption was made that the travel time distribution was known. Furthermore the marginal cost of early and late arrival were also known to the traveller. Based upon these conditions, Gaver showed it would be possible to predict the travellers' optimal departure time; the value of departure time that minimised the total expected cost of earliness and lateness combined (or alternatively, maximised the EU). An assumption of

Gaver's was that travellers' primary behavioural response to travel time risk is to modify their departure time. However this has proven to be a foundation of much subsequent work in the field (e.g. Polak 1987a, Bates et al, 2001). It should be noted that Gaver also discussed route choice as an alternative behavioural response, but did not apply this in his modelling framework. Gaver formally derived the extra amount of travel time allowed in response to travel time risk, which was named the *safety margin* or *headstart*. A relationship between the safety margin, risk aversion and travel time variation was established. This contribution of Polak's will be particularly useful when the subject of *reliability equivalence* is introduced in the last part of this section.

Other early work in this area includes Vickrey (1969) which was seemingly developed independently of Gaver. Vickrey's focus was upon areas of the highway network where demand exceeded supply; so-called bottlenecks. It was recognised that these bottlenecks would create queuing conditions which would therefore cause individuals to arrive after a desired time. The existence of a desired arrival time in Vickrey's model essentially supported Gaver's PAT. Vickrey's model explicitly placed a value on remaining at home prior to embarking upon a trip, as well as different costs incurred by the traveller for early arrival and late arrival at their destination, with lateness incurring higher costs (this idea has more recently been re-examined by Fosgerau and Engelson, 2011). Vickrey's model forms only a small section of the overall paper, the majority being an argument for tolls as a method for addressing congestion. Therefore it is left to later work to formalise this approach into the Scheduling framework, as it is known today.

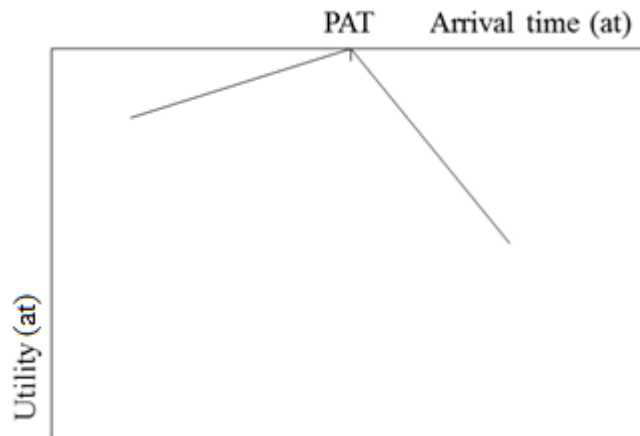
Knight (1974) perhaps owes more to Gaver than Vickrey. Knight's work was an attempt to explain how improvements in reliability would be of benefit by reducing a traveller's safety margin. It was effectively an attempt to formalise Gaver's approach. A contribution that is rarely attributed to Knight is an acknowledgement of the shape of the utility function (with trip duration on the  $x$ -axis); this was the first acknowledgement in transportation that the curvature of the function that would indicate a traveller's perception of risk. Knight is useful to the context of this thesis in that it takes into account public transport (PT) operations specifically: that the existence of a timetable and staffing issues on PT will result in the impact of unreliability being felt more

acutely than on private modes. Despite the initial focus on PT, Knight restricted use of this version of the Scheduling model to situations where departure times could be varied continuously. Furthermore, it was assumed that the commuter does not adjust this departure time, but rather accrues a greater safety margin as a result of improvement to reliability. This should be considered a regressive development from Gaver. In the latter sections of Knight (1974) it was concluded that the safety-margin based benefits of reliability could in some circumstances be greater for public transport uses than for the private car, but this fails to take into account Gaver (1968) (and much that has come since), who found that an optimal departure time would be unlikely to be achieved by public transport users.

This apparent confusion and overlap between three of the key early works in the development of the Scheduling approach therefore required some clarification. The development of the Scheduling approach as it is used today is usually attributed to Small (1982). This work was primarily empirical in nature, but nevertheless focussed upon the Scheduling utility function. It was an attempt to explain the changing characteristics of peak demand which would be determined by departure time. Small's work can be linked to earlier work by a reliance on the concept of the PAT and the optimisation of departure times. If the effect of demographic and other traveller characteristics were ignored, and travel time was fixed, the Scheduling model proposed and estimated by Small is:

$$U(at) = \beta \cdot SDE + \gamma \cdot SDL \quad (2.16)$$

Where  $U(at)$  is the utility at arrival time  $at$ .  $SDE$  represented the duration of time prior to the PAT, and  $SDL$  is the temporal duration after the PAT. The parameters  $\beta$  and  $\gamma$  are the marginal utilities of  $SDE$  and  $SDL$  respectively. Equation 2.16 is represented in Figure 2.4.



**Figure 2.4** - A Scheduling utility function based upon Small (1982) for a fixed travel time

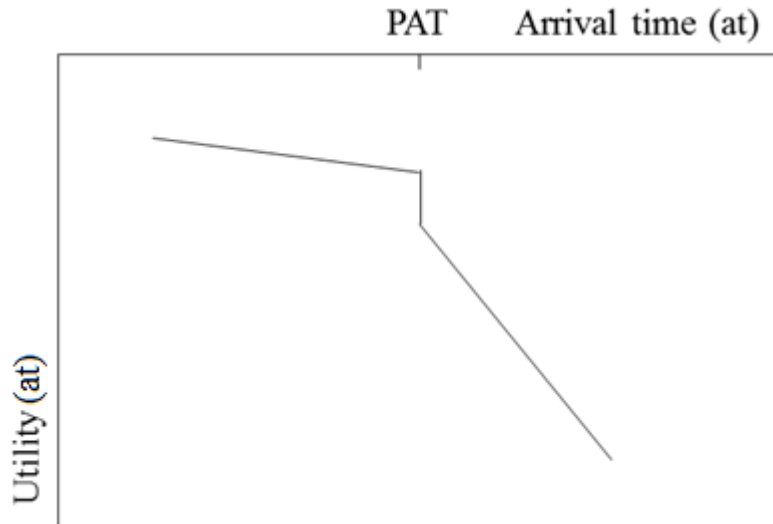
Moving from left to right on the  $x$ -axis, Figure 2.4 shows the amount of negative Scheduling utility (disutility) decreasing as the function approaches the PAT. At an arrival time equal to the PAT, no Scheduling disutility is incurred. The utility then decreases at a rate of  $\gamma$  for arrival times after the PAT. Evidence was provided by Small (1982) that travellers dislike lateness more than earliness:  $\beta > \gamma$ . Small also suggested a fixed penalty would exist for lateness, so that as soon as the traveller was late at their work destination, a further fixed amount of disutility was incurred. This is analogous to losing pay or reputation for late arrival at work. This lateness penalty, combined with the value of time, resulted in the common Scheduling approach as commonly found in the literature (Bates et al, 2001).

$$U(\text{departure time}) = \alpha \cdot t + \beta \cdot SDE + \gamma \cdot SDL + \theta D_L \quad (2.17)$$

Where  $\alpha$  is the marginal utility of travel time,  $t$  the travel time,  $D_L$  is a dummy variable for lateness, taking a value of '1' if late and '0' otherwise, and  $\theta$  is the value of (dis)utility associated with the lateness dummy variable. All parameters are expected to be negative.

If the assumption is made that  $\beta > \alpha > \gamma$ , then the utility function can be re-plotted to incorporate travel time and the lateness penalty in combination. This is done in Figure 2.5, under the assumption of a fixed departure time. The slope of the utility function

(with respect to  $t$ ) in earliness is given by  $(\alpha - \beta)$  and the slope in lateness is given by  $(\alpha + \gamma)$  (Batley 2007).



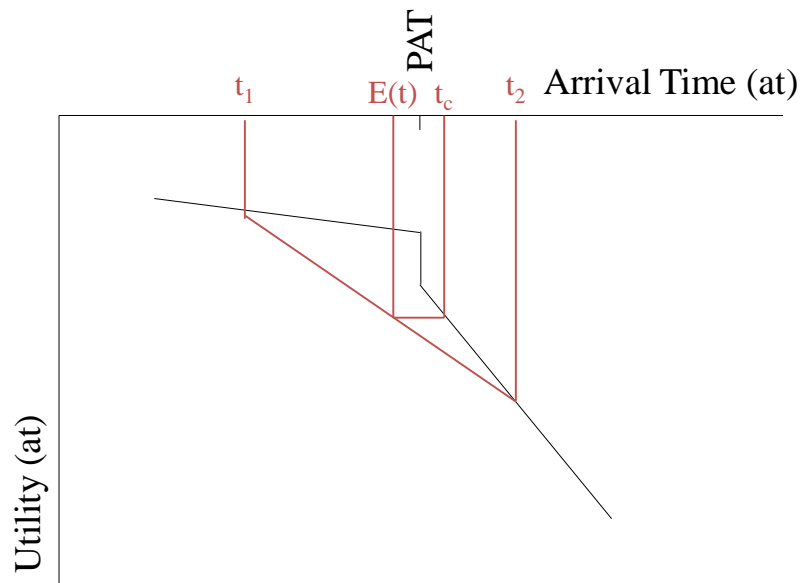
**Figure 2.5** - A Scheduling utility function based upon Small (1982) and Batley (2007)

The issue with the framework shown in Equation 2.17 is that it does not take into account reliability concerns i.e. it is a utility function rather than an expected utility function. Pells (1987) utilised the Scheduling approach to demonstrate that uncertainty in travel time would cause the traveller to allocate a greater amount of time to travel. Therefore reliability could be valued in terms of opportunity cost of that extra time. The aforementioned work of Polak (1987a) also dealt with reliability and earliness and lateness, but did not develop Equation 2.17 into an expected utility (EU) context where the overall travel times were less certain. This development was made by Noland and Small (1995). Noland and Small recognised that total travel time could be divided into three categories: free flow time, recurrent congestion and random travel time. This was a key insight: reliability could now be defined as the random element of travel time that was unpredictable. Noland and Small showed that for all distributions of this random element, EU would be maximised as follows:

$$EU(\text{departure time}) = \alpha \cdot E(t) + \beta \cdot E(SDE) + \gamma \cdot E(SDL) + \theta P_L \quad (2.18)$$

In Equation 2.18, expectations are taken of  $t$ ,  $SDE$  and  $SDL$ . The dummy variable of lateness from Equation 2.17 becomes  $P_L$ , which is now the probability of lateness. Noland and Small used the same parameters in both 2.17 and 2.18.

Batley (2007) showed how the EU function of Equation 2.18 could be used to represent risk attitudes through the *risk premium* as was previously shown for the case of MV in Section 2.1.



**Figure 2.6** - A Scheduling utility function based upon Noland and Small (1995) and Batley (2007)

Figure 2.6, originally proposed in this form by Batley (2007), demonstrated the existence of risk aversion through the Scheduling utility function for a fixed departure time. Similar to Figure 2.2,  $E(t)$  represents the expected value of  $t$ ,  $t_c$  is the certainty equivalent, and the difference between these values is the ‘reliability premium’.  $t_1$  and  $t_2$  represent risky outcomes. The angle formed between linear sections of the function was shown to create concavity and thus imply risk aversion. This suggested further similarity with the MV utility function shown in Section 2.1 and is a useful result in advance of the theory of reliability equivalence that will be outlined in Section 2.5. The latter part of Chapter 6 will investigate whether these reliability premia for MV and Scheduling utility functions are in fact similar.

### 2.3.2 The Mean-Lateness Approach

The *Mean-Lateness* approach is a framework for approaching transport reliability that is extensively used in the UK rail industry. The approach is defined in the Passenger Demand Forecasting Handbook, although as PDFH is not in the public domain this introduction is based upon the interpretation provided by Wardman and Batley (2014) and Batley and Ibáñez (2012). The approach is specific to public transport operations insofar as the PAT of the Scheduling approach is replaced by the timetabled arrival time at the destination. If the travel time is greater than the scheduled travel time, then lateness is incurred

Unlike the aforementioned MV and Scheduling frameworks, the approach is used primarily to predict passenger demand responses to changes in Mean-Lateness: i.e. an elasticity of demand with respect to reliability. The basic utility function for service  $i$  takes the form:

$$U_i = \alpha \cdot AML_i + \beta \cdot ST_i + \dots \quad (2.19)$$

where AML is the average minutes lateness (compared to the scheduled arrival at the destination) and ST is the scheduled travel time. Therefore the ratio of alpha/beta is the value of Mean-Lateness in units of scheduled time. This ratio is referred to as the lateness multiplier in the PDFH (ATOC, 2013).

Mean-Lateness has found use within the UK's rail industry as it can be integrated with existing guidance within the PDFH. In their review, Wardman and Batley (2014) outlined how the PDFH generalised cost function (traditionally comprising of travel time, headway and interchanges) could be supplemented by the Mean-Lateness multiplier to predict a change in demand

$$\frac{V_{new}}{V_{base}} = \left[ 1 + \frac{\omega_{AML}(L_{new} - L_{base})}{GJT_{base}} \right]^{\eta_{GJT}} \quad (2.20)$$

Where  $V_{new}$  is the volume of demand in the new situation;  $V_{base}$  is the volume of demand in the base situation.  $L_{new}$  and  $L_{base}$  are Mean-Lateness in the new and base situations respectively, and  $\omega_{AML}$  is the valuation of this lateness in units of scheduled travel time.  $\eta_{GJT}$  is the elasticity of demand with respect to GJT. Wardman and Batley



(2014) showed that an implied lateness elasticity ( $\eta_{AML}$ ) can be calculated from the above using:

$$\eta_{AML} = \eta_{GJT} \frac{\omega_{AML} AML}{GJT} \quad (2.21)$$

Estimates of  $\omega_{AML}$  have often been made using SP data. Another strand of research has attempted to estimate the impact of reliability on demand using the Punctuality Performance Measure (PPM). The PPM is the proportion of short distance trains arriving at their destination within 5 minutes of the scheduled time (or long distance arriving within 10 minutes). Wardman and Batley (2014) showed how the elasticity of demand with respect to PPM could be used to calculate  $\eta_{AML}$ :

$$\eta_{AML} = \eta_{PPM} \frac{\delta PPM}{\delta AML} \frac{AML}{PPM} \quad (2.22)$$

Using Equations 2.21 and 2.22 it is possible to obtain and compare estimates of  $\eta_{AML}$  from both SP and RP experiments. Wardman and Batley (2014) did this, and such analysis will be useful to this thesis (and will form part of the discussion in 2.4.1 and 2.4.2 of the present chapter).

The Mean-Lateness approach overcomes a key issue of the Scheduling approach that it is often difficult to obtain a passenger's PAT without asking them directly. It is also of practical use insofar as it can be applied to station to station flows (Wardman and Batley, 2014). It does however imply no negative value for early arrival, and the form in 2.19 does not take into account departure time at the origin station. The standard ML model was extended by Batley (2007) to account for the latter.

### 2.3.3 Reliability Equivalence

The three frameworks outlined in the previous section are all effectively attempts at representing the effect of reliability on travellers. It logically follows that some correspondence exists between them. The literature has been particularly interested in the correspondence between the MV and Scheduling frameworks, and theoretical equivalence has been established between them. This equivalence is referred to as *the theory of equivalence* or *reliability equivalence* for the remainder of this thesis.

In Section 2.3.1 it was shown that Noland and Small (1995) formally translated the Scheduling approach from a context of travel time certainty to one involving travel time risk. This development therefore introduced a second EU function for reliability, alongside MV. An open question in the literature is: which of the two should be chosen for use in the valuation of reliability – MV or Scheduling under risk?

If it could be proved that MV and Scheduling EU functions were equivalent, it would have the impact of removing the choice between the two frameworks: each could be considered equally relevant and could be chosen according to surveying or other practical constraints. The first key step toward this end was made by Gaver (1968) who derived the optimal probability of lateness from Scheduling parameters only.

Bates et al (2001) recognised that all travellers who accord with expected utility theory would accept some value of probability of lateness greater than zero. Previous work had established that there was a positive relationship between the level of unreliability and the headstart (Pells, 1987). ‘Headstart’ (as defined by Gaver, 1968), was shown by Bates et al (2001) to be approximately linear in standard deviation, such that the optimal Scheduling EU is a linear function of mean and variance. This is the key result of Bates et al (2001):  $\beta E(SDE) + \gamma E(SDL)$ , derived by Noland and Small, could be shown to be approximately equal to  $\rho\sigma$  when departure time was optimal. This was demonstrated in Bates et al (2001) where the Reliability Ratio was reformulated to consist of Scheduling parameters only.

Bates et al (2001) went on to consider discrete departures, such as those that might be experienced on public transport services, and showed that this equivalence no longer held in most cases. However this is less clear when the traveller is ignorant of timetabled departure times and arrives at the departure point randomly.

Fosgerau and Karlström (2010) built upon Noland and Small (1995) and Bates et al (2001) to provide mathematical proof of the equivalence between MV and Scheduling. They specified the EU function as:

$$EU = \max_D EU(D, T) = \max_D \left[ \beta D + (\alpha - \beta)\mu + (\beta + \gamma) \int_{\frac{D-\mu}{\sigma}}^{\infty} (\mu + \sigma t - D)\phi(t)dt \right] \quad (2.23)$$

Where

$D$  is the departure time

$\alpha, \beta, \gamma$  are the Scheduling parameters defined by Small (1982)

$t$  is a standardised random variable, and  $\mu$  and  $\sigma$  represent the mean and standard deviation of the travel time distribution respectively

Given that  $t$  is a random variable which could be drawn from any distribution, it is possible to derive an expected utility function in the form of MV but using Scheduling parameters (from Bates et al, 2001) for any distribution:

$$EU = \alpha\mu + (\beta + \gamma)H\left(\Phi, \frac{\beta}{\beta + \gamma}\right)\sigma \quad (2.24)$$

Where  $H$  is calculated from a standardised travel distribution ( $\Phi$ ) and the Scheduling parameters, as follows.

$$H\left(\frac{\beta}{\beta + \gamma}\right) = \int_{1 - \frac{\beta}{\beta + \gamma}}^1 (s) ds. \quad (2.25)$$

The marginal expected utility of standard deviation from the MV approach is therefore estimated by  $H$  and the Scheduling parameters, which implies equivalence between Scheduling and MV subject to the constraints outlined. Theoretically, this is a powerful result which provides clarity in the field as to which is the preferred framework to use. There are some limitations to this however. The departure time needs to be a continuous variable, which excludes the equivalence result from public transport applications. It should be noted however that subsequent work has overcome this issue; Fosgerau and Engelson (2011) defined time varying utility rates at the origin and destination and showed that the influence of travel time variable would be the same irrespective of whether the transport mode was scheduled or not. Another practical issue relating to the equivalence result is that to convert MV parameters to Scheduling parameters, the optimal probability of lateness is also required (which would not always be available). This is due to the relationship (from Equation 2.24) of:

$$\rho = (\beta + \gamma)H \quad (2.26)$$

Where it is not possible to separate the elements of  $(\beta + \gamma)$  unless the term  $\frac{\beta}{\beta + \gamma}$  is also known.

Despite such practical issues, studies have attempted to find empirical evidence for this equivalence, which will be covered in Section 2.4.3.

## **2.4 Empirical Application of Reliability Frameworks**

A substantive amount of research in this field has been focussed towards eliciting a value of reliability for use in investment appraisal. Research has made use of MV, Scheduling and Mean-Lateness approaches in this endeavour, but no standard has emerged. Methodologically, there has been a general trend to favour *Stated Preference* (SP), where respondents indicate their favoured choice from two (or more) hypothetical options. This section will begin by outlining some of the key SP studies and their contribution towards improving the methodologies for estimating a value of reliability (VOR). The next part will look at SP valuations of reliability in light of these methodological issues. The final part of the section on the SP methodology will bring together these results and methodological issues and highlight some general notes of caution of using SP in this context.

The primary alternative to SP is Revealed Preference (RP), where consumer behaviour in the marketplace is observed directly. There is a small literature related to the estimation of a VOR using RP studies which will be outlined. These include a literature on the estimation of late time multipliers using Mean-Lateness (ML) models. Based on this review of RP evidence some concerns with that approach will also be highlighted. The section will conclude with a discussion which will compare the merits of the two approaches and make the case for the use of RP.

### **2.4.1 Stated Preference**

This section will outline the studies that have taken place using SP; the first part will focus on those that have made a methodological contribution towards estimating an SP VOR. In the second part of Section 2.4.1 the results arising from SP studies will be outlined. The final part of this SP section will critique the widespread use of SP in VOR studies.

## Methodological Development of SP Reliability Studies

The previously mentioned work of Jackson and Jucker (1982) deserves credit for being the first to apply MV in a transportation context. After presenting the model (of Equation 2.9), the authors estimated the distribution of  $\lambda$  based upon an SP experiment. In the course of this exercise, the authors identified two practical issues related to SP experiments in the field of reliability. The first being the correct method of presenting travel time risk to respondents, and the second being the form of the MV expected utility function – in particular the indicators used.

The complexity of presenting reliability to survey respondents is recognised in much of the SP research. Senna (1994) provides an overview of early attempts to present TTV. Senna made two adjustments to his own survey in light of his review: that one of the options presented should contain zero variability, and that choice situations should be related to travellers' actual experiences. This latter point was of particular foresight in light of the work that has come since (e.g. Hensher, 2010).

Jackson and Jucker (1982) have not been the only authors to utilise the standard MV EU function of Equation 2.8. Some authors have been concerned that presenting a Normal distribution of travel times to respondents might be misleading and have therefore utilised the median and other percentile based measures of spread (Small et al, 2005). One example of such a measure is the reliability buffer time, which is the difference between a higher percentile of travel time (most commonly the 80<sup>th</sup> or 90<sup>th</sup> percentile) and the median – thereby measuring the width of the right side of the distribution. Those that have adopted this measure (Small et al, 1999; Small et al, 2005; de Jong et al, 2009) have suggested that these percentile based measures should be preferred to the standard deviation/variance. However, other authors have found that the mean and variance provide better model fit using their SP data (Hollander, 2006). This question of the correct form of the MV equation remains a live issue and is covered in greater detail in subsequent chapters.

The Scheduling approach (of Equation 2.18) has also often been modified in empirical research. Small et al (1995) identified correlation between  $SDL$  and  $P_L$  within their model. It was found that the fit of their model was improved if  $SDL$  was omitted

from estimation; effectively that  $P_L$  was capable of accounting for the entire impact of lateness. Small et al (1999) estimated a quadratic Scheduling model, where parameters for  $SDE^2$  and  $SDL^2$  were estimated alongside the standard ones. The parameter for  $SDL^2$  was found to be insignificant, but the parameter associated with  $SDE^2$  was significant and could potentially have an impact on the shape of the Scheduling utility function. However Hollander (2006) did not find this parameter to be significant in his own model.

Another research strand has been to include demographic variables in SP choice models. Pells (1987) was one of the first to do so, and found that seniority in the workplace had an impact; more senior employees had lower values of reliability; presumably as they exercised greater control over their working day. Small et al (1995) conducted an SP where respondents were asked to give information related to their sex, age, income, and whether they had children. The most detailed part of this work was focussed on employment status, where individuals were asked about their conditions of employment, whether they were self-employed, and finally their income. Contrary to Pells, income did not appear to affect the valuation of Scheduling parameters. It was found that, in general, salaried workers were less averse to lateness than wage earners, although this was not an unexpected result. Small et al (1995) also found that commuters who had parenting responsibilities tended to have higher values of time and reliability. This finding may go some way to explaining different attitudes toward time and reliability between males and females that have been found in the literature (Brownstone and Small, 2005), assuming that there is a systematic difference between sexes in terms of responsibility for childcare. Asensio and Matas (2008) similarly accounted for income in their SP experiment but also found no effect of this variable on the valuation of time or reliability. The explanation given was that only a small range of incomes were captured in the sample.

One concern running throughout the SP literature is the best method of conveying travel time variability to respondents who are not necessarily familiar with statistical indicators such as the standard deviation. Senna (1994) is an example of the most commonly used method, where two options were presented to the respondent, each with five equiprobable travel times and a mean. From these five values it was presumed that

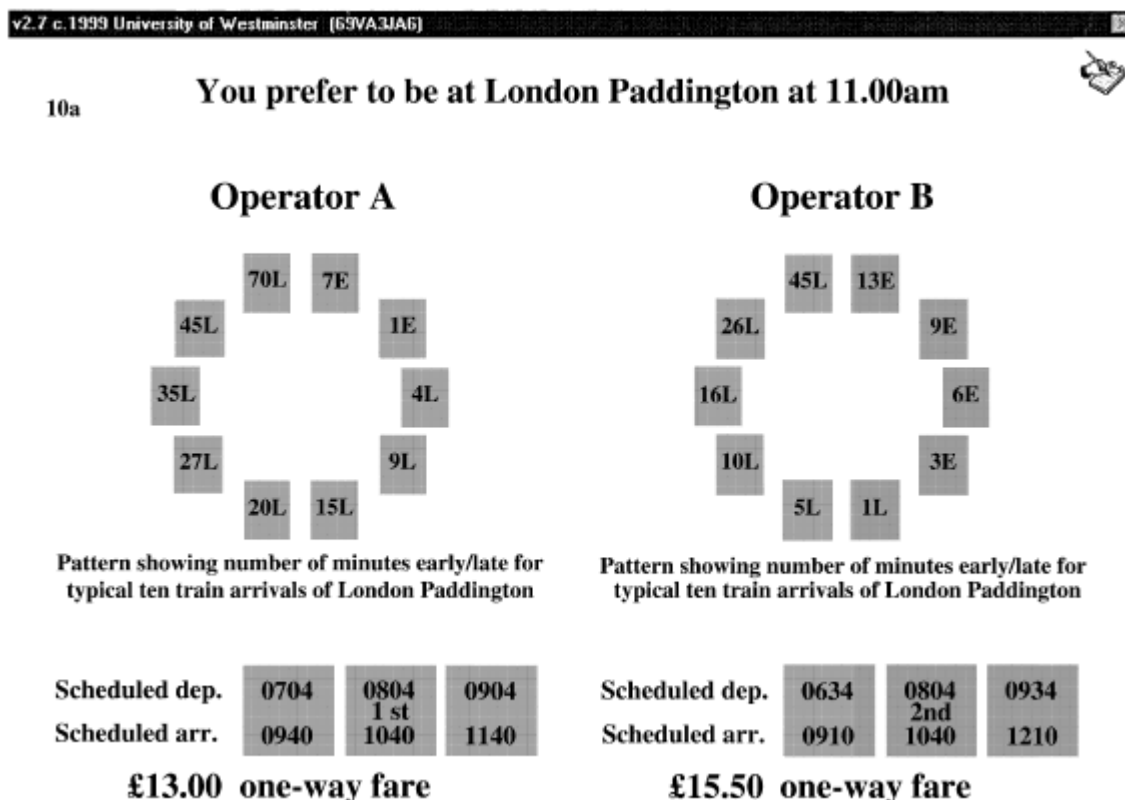
a proxy for the standard deviation could be estimated by the respondent. Small et al (1999) took a similar approach, although they also considered presenting ten travel times for each choice option so that a travel time distribution could be more accurately represented. It was concluded that this approach may overload the survey participant with information and lead to fatigue with the questionnaire.

Benwell and Black (1984) did however present ten travel times for each option. As part of this presentation, the delays were ordered so that the highest delays were on the right hand side. This approach has been questioned in the literature, since some respondents might interpret this as implying a deteriorating quality of service (Bates et al, 2001). A further concern of Bates et al was that for some options, a substantial proportion of the travel time delays were zero – resulting in infrequent travellers perhaps assuming that delay would not affect them. In response to these concerns, Bates et al (2001) used a clock face presentation (shown in Figure 2.7), where varying levels of delay were presented in a circular fashion to reinforce that each event was of equal probability. A test was given to participants to check their understanding of the method. Those that poorly understood the concept were excluded from analysis. Whilst this process appeared sensible in the context of an SP, it is possible that there is a proportion of the population in real situations who would struggle to understand reliability as a concept. Perhaps the exclusion of some groups from SP might bias parameter estimates away from their true population values.

The work of Tseng et al (2009) acknowledged this concern by testing how well participants understood a variety of the SP presentational formats mentioned above (as well as a vertical bar format of Hollander, 2006 and Batley and Ibáñez, 2009; shown in Figure 2.8). The starting point of that paper was a concern that participants were not fully comprehending surveys on reliability. To test this hypothesis, Tseng et al split their sample of survey respondents into ‘lower’ and ‘higher’ educated groups. They found that questions about reliability were better answered by the higher educated group, but this was not true in all cases. A key finding of Tseng et al was that participants indicated a preference for the format of five travel times making up each choice alternative, and furthermore the questionnaire would be administered in person by survey staff. They also indicated preference for the vertical bar presentation of

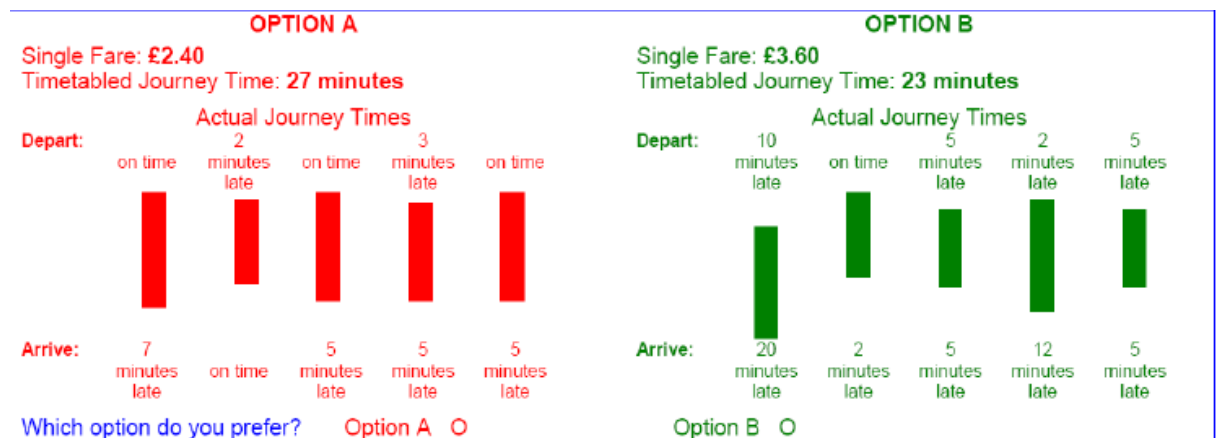
Hollander (2006). Tseng et al noted from their experience that it was hard to recruit participants for their survey, which meant that they offered a small amount of money to participants upon completion of the survey. Tseng et al were aware that offering a financial incentive might mean that their sample would be biased towards respondents with a lower value of time or reliability, therefore resulting in lower estimates of these values.

They concluded by saying that potential existed for survey respondents to misapprehend both the survey and the reliability issues presented to them in an SP context, and great care would be required to overcome such issues. Furthermore, although they concluded that researchers should simplify their surveys to present five travel times of equal probability, this had the potential to misrepresent the nature of reliability – for example by not adequately representing long travel times with low probabilities which are of great concern to many travellers.



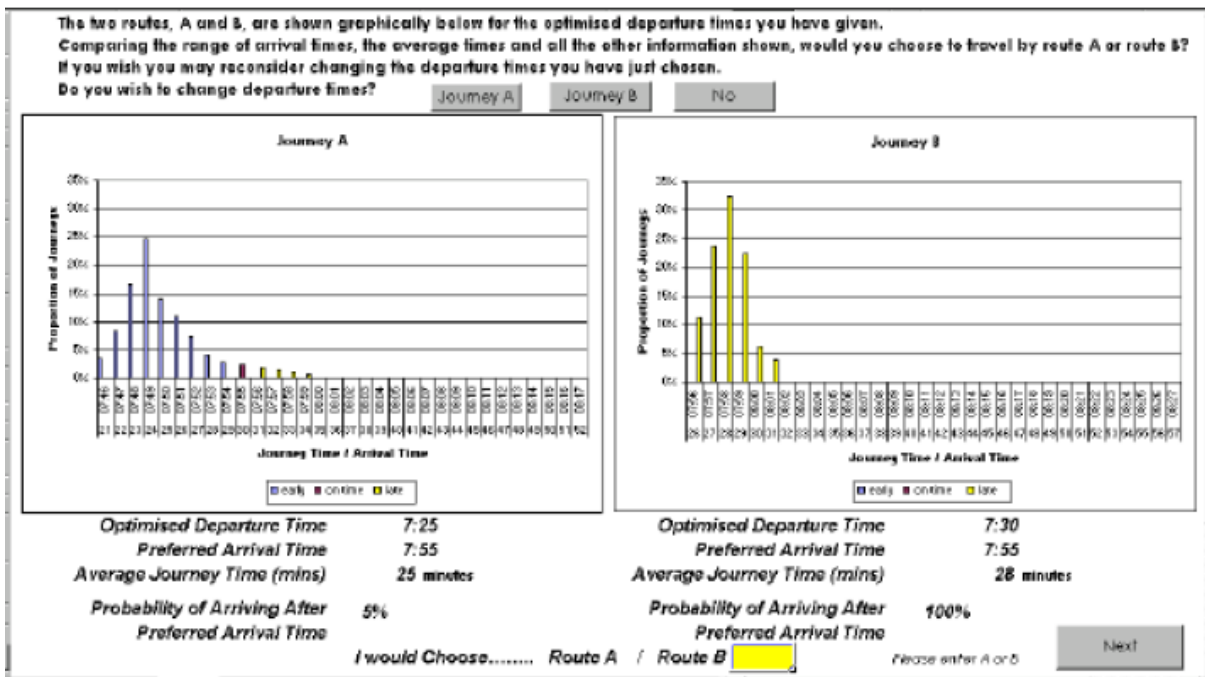
**Figure 2.7** - The 'clock face' presentation of an SP survey as used by Bates et al (2001)





**Figure 2.8** - The vertical bars presentation of an SP survey as used by Hollander (2006) and Batley and Ibañez (2009)

The final method of presentation, proposed by Copley et al (2002), was to present a travel time distribution to survey participants (shown in Figure 2.9). This was contrary to the majority of the reliability SP literature where it has been assumed that such detail would overburden the participants. Tseng et al showed that this method of presentation is one of the least favoured in their study. Nevertheless, Copley et al (2002) detail how participants could be encouraged to understand a travel time histogram or distribution with support from those conducting the survey. Not only did this allow greater detail to be conveyed, but allowed innovative surveying techniques such as allowing participants to optimise their departure time dynamically based upon the travel time distribution.



**Figure 2.9** – The travel time distribution based presentation of SP as used by Copley et al (2002)

The above review provides a range of methods for conducting SP surveys in a reliability context. Presentation of reliability is clearly a subject of concern and the subject of academic debate in itself; the reviews of Bates et al, 2001; Tseng et al, 2009, Li et al, 2010 and Wardman and Batley, 2014 are testament to this. The issue that these reviews have attempted to address is the difficulty in conveying what is a complex subject to members of the general public. It is clear that misapprehension of the survey method itself is a problem, and one which has not been fully addressed in an SP context. Tseng et al found that great care was required in choosing the format and administering SP questionnaires, and there was potential to obtain responses based upon misapprehension on the part of the respondent given the complexity of reliability issues.

### Stated Preference VOR Estimates

The preceding discussion was focussed upon the methodological issues of estimating an SP VOR. In this section, the VORs estimated using SP will be the focus. Studies will be introduced in chronological order and where possible the Reliability Ratio will be presented. As this section focuses upon Reliability Ratios, it will consist primarily of

studies that utilised the MV and Scheduling frameworks; Bates et al's (2001) formula for calculating a Reliability Ratio from scheduling parameters will be utilised:

$$RR = \frac{\beta}{\alpha} \ln \left( 1 + \frac{\gamma}{\beta} \right) \quad (2.27)$$

Where RR is the Reliability Ratio, and  $\alpha, \beta, \gamma$  were defined in Equation 2.18 of this chapter.

Black and Towriss (1993) were the first researchers to present the 'Reliability Ratio'. It is defined as the ratio of the marginal utility of reliability (usually measured in units of standard deviation) to the marginal utility of travel time. The Reliability Ratio they presented for car based commuters, 0.55, would appear low in relation to much of what has come since, although more recent evidence from the UK actually puts the average value at 0.4 (ITS and Accent, 2015). Black and Towriss' results also suggest that the RR would be lower for bus users than for rail users.

The respondents to the study by Tseng et al (2009) indicated that the most preferred questionnaire format was that of Small et al (1999). This would imply that the central Reliability Ratio estimate of 2.51 of Small et al (1999) is worthy of consideration although it is high in relation to many other studies. This questionnaire format is only slightly different to that used by Black and Towriss. Where Small et al's contribution differed substantively was in their use of demographic data interacted with travel time and reliability. Small et al (1999) was actually building upon the work of Koskenoja (1996), who was the first to recognise the value of demographic data in this area. Koskenoja found that income and whether the traveller had young children would affect the result. The Reliability Ratios estimated were closer to one, with a mean average of 0.77.

Hensher (2001) divided travel time into a number of components: free-flow, slowed down, stop/start and uncertainty and assigned different values to each of these categories to represent alternatives. It could be reasonably argued that travellers would not explicitly distinguish these components of reliability. Therefore a low Reliability Ratio might be expected, associated with respondent misunderstanding of the questions. The RR estimated of Hensher's favoured model was 0.67.

Hollander (2006) is an interesting and often quoted case due to its low reported RR of 0.1. Although the methodology presented by Hollander was not contentious in relation to the rest of the literature, there were sampling issues that may have affected the results. For example, Hollander drew his sample entirely from bus users, and conducted the SP online. This could have led to a biased sample (for example by over-representing young, low income workers). Kouwenhoven et al (2014), for example, found that an internet only survey resulted in lower VOT estimates than expected. Moreover, Hollander's sample only covered a single geographic area, and it is difficult to know how reliability levels in that area might have affected respondent attitudes.

Asensio and Matas (2008) estimated a RR of 0.98 using a presentation format similar to Small et al (1999). Further support for use of the presentational format of Small et al was found by Li et al (2010) who also estimated an overall Reliability Ratio of 0.7. Li et al also suggested that commuters should value travel time and reliability more highly than non-commuters due to the existence of the PAT, but it is non-commuters who had the larger RR in their experiment.

The SP conducted by de Jong et al (2009) focused on road freight transport and estimated a RR of 1.2, although this was later acknowledged as a provisional result in de Jong et al (2014), where water-freight based Reliability Ratios of between 0.1 and 0.4 were estimated.

Tilahun and Levinson (2010) supported the SP presentation of Copley et al (2002), who favoured presenting histograms to travel time to respondents. They calculated a RR of 0.89, which is similar to many of the studies already outlined. Nevertheless, the use of histograms to represent a travel time distribution in SP experiments remains unpopular.

Kouwenhoven et al (2014) summarise the most recent major Dutch VOT and VOR study, taking into account a range of modes and demographic characteristics. They presented five equiprobable travel times to respondents. They noted that those willing to undertake their SP study as part of an internet panel had a lower VOT than had been expected, and so additional data collection was required. Although the study quotes a range of RRs, their main estimate is 0.7 (for both business and other purposes).

Similarly, Ehreke et al (2015) summarise the reliability element of the most recent major German VOT study. They found that among car drivers, WTP was higher for time savings than reliability savings, and estimated a RR of 0.7 for this group.

The final strand of SP evidence to present are estimates of  $\omega_{AML}$  used in PDFH (based upon the summary provided by Wardman and and Batley, 2014). The SP conducted by Benwell and Black (1984) estimated a  $\omega_{AML}$  of 2.5, which was used in the first edition of PDFH. This value remained constant in the second edition of PDFH, with the exception of commuting trips being reduced to 1.25 based upon evidence provided by MVA (1989). PDFH 3 reverted back to a  $\omega_{AML}$  of 2.5 for all purposes, whilst PDFH 4, taking into account Bates et al (2001), increased the  $\omega_{AML}$  to 3. In addition to the values in PDFH, Batley et al (2006) estimated a central  $\omega_{AML}$  of 2.7, although this varied from 1.8 to 5.3 depending on the trip type. Faber Maunsell and Mott Macdonald (2007) estimated a  $\omega_{AML}$  of 5.3, which is not dissimilar from the value calculated by MVA and ITS (2012) for business travellers. Lu et al (2008) estimated a value of 3.27 for this same group of travellers. Wardman and Batley (2014) also reported a number of unpublished studies including TPA (1992) (a  $\omega_{AML}$  of 1.7) and MVA (2000) (a  $\omega_{AML}$  of 5.7).

Wardman and Batley calculated an implied  $\eta_{AML}$  (using Equation 2.20) taking into account the highest and lowest recommended values of  $\eta_{GJT}$  within PDFH 5.1. The travel with the highest  $\eta_{GJT}$  was London long distance flows (L-LD) where a value of -1.35 was recommended. Taken together with a Mean-Lateness of 5.2 and an assumed GJT of 120 minutes, Wardman and Batley estimate an implied  $\eta_{AML}$  of -0.18. The lowest  $\eta_{GJT}$  recommended by PDFH 5.1 was for non-commuting travel within London (L-NC) (a value of -0.9). Under the assumptions of a Mean-Lateness value of 1.3 and a typical journey time of 30 minutes, the implied  $\eta_{AML}$  was calculated as -0.11. Based upon the above evidence and the methodology of Wardman and Batley, it is possible to calculate an implied  $\eta_{AML}$  for the range of SP studies where a  $\omega_{AML}$  was estimated.

**Table 2.1** - Implied ML elasticities derived from SP estimates of the value of Mean-Lateness

	$\omega_{AML}$	$\eta_{AML}$ L-LD	$\eta_{AML}$ L-NC
Benwell and Black (1984)	2.50	-0.15	-0.10
MVA (1989)	1.25	-0.07	-0.05
TPA (1992)	1.70	-0.10	-0.07
MVA (2000)	5.70	-0.33	-0.22
Bates et al (2001)	3.00	-0.18	-0.12
Batley et al (2006)	2.70	-0.16	-0.11
Faber Maunsell/Mott Macdonald (2007)	5.30	-0.31	-0.21
Lu et al (2008)	3.27	-0.19	-0.13
<b>Mean Average</b>	<b>3.18</b>	<b>-0.19</b>	<b>-0.12</b>
<b>Standard Deviation</b>	<b>1.58</b>	<b>0.09</b>	<b>0.06</b>
<b>Coefficient of variation</b>	<b>0.50</b>	<b>-0.50</b>	<b>-0.50</b>

In Table 2.1, L-LD (London-Long Distance) refers to the situation where  $\eta_{GJT}$  is equal to -1.35, with a GJT of 120 minutes (where GJT is made up of scheduled time, headway and number of interchanges), and a mean-lateness of 5.2 minutes. L-NC (London- Non Commuting) refers to the situation where these values are -0.9, 30 minutes and 1.2 respectively.

The summary statistics calculated in Table 2.1 (mean, standard deviation, coefficient of variation) will be used later in this chapter as a comparator to RP based studies.

### Discussion of SP methodology and VOR estimates

Whilst issues around respondent understanding of reliability can be at least partially ameliorated by careful presentation and design of surveys, there has been other work that has suggested that there are further underlying problems with SP in reliability (and travel time) contexts that must be taken into account.

A critique of SP that can be made generally is that the hypothetical answers given by survey participants would not reflect the choice they would make in reality. In a general transport context, Wardman (1988) provided possible reasons for this discrepancy on the part of the respondents:

- Misapprehension of the survey and its contents.
- Not giving the survey adequate attention; including fatigue and boredom (Bradley and Daly 1994, although the importance of this effect was later questioned by Hess et al, 2012).
- Inertia: favouring the same type of option and not trading between attributes.
- Protest response where individuals consciously seek to influence the outcome of the survey.

Wardman also acknowledged that apparent non-trading or protest responses may actually be a reflection of a high/low value of time outside the scope of the survey. A study specific to rail and reliability issues suggested that the key reason for the divergence between hypothetical and actual behaviour was the protest response (Wardman and Batley, 2014). The evidence for this was that divergence between RP and SP valuations would decrease when the purpose of the SP was obscured from the participants. The study also highlighted the need for good RP evidence as a comparator for SP results. Brownstone and Small (2005) compared results of an SP and RP study based upon car users with a choice between freeway and tolled alternatives. In addition to the issues above, they highlighted a number of differences between the two methodologies in this context:

- Their RP estimates of VOR and VOT were substantively lower than the corresponding SP valuations.
- That travellers may intend on using a slower/more unreliable route, but on a given day circumstances may force them to use the faster/reliable tolled route.
- Travellers were likely to overestimate the impact of experienced congestion, which translates to a misinterpretation of the SP.

Hensher (2010) focused upon the existence of ‘cheap talk’ in his analysis of the divergence between RP and SP. Cheap talk in this context being the respondent not taking the survey seriously (as suggested by Wardman, 1988) as well as not having the same rewards and disincentives as one would find in a real-world situation (Lu et al, 2008). Hensher made some suggestions which will improve the results from SP:

- Recruit participants that are knowledgeable about the transport context in which the survey is set; they will take it more seriously.
- Include an option for the participants to ‘opt-out’ where they are indifferent or unsure about their preferences.
- Include additional questions to ascertain the choice strategy employed by each participant.
- In the context of toll experiments, offer participants a voucher/incentive to use an actual toll road so that they have real experience of both options.

The development of these suggestions appears to offer a defence of SP as a method for the valuation of reliability. Other studies have however suggested that RP would be preferred if the data were available (e.g. Bates et al, 2001).

### **2.4.2 Revealed Preference**

This section will begin with an exploration of the methodological development of Revealed Preference (RP) studies before developing onto the VOR estimates made using the technique. This section will bring together a summary of SP and RP VOR studies in Table 2.3 and conclude with a comparison of the two techniques.

#### **Methodological Development**

Revealed Preference (RP) methods are less utilised than SP VOR contexts. Hensher (2010) provides explanation for this by suggesting that there remain significant issues with RP:

- Un-chosen alternatives are often imposed upon individuals’ choice sets which they may not have been aware of. One could add that it would often be problematic to obtain a full choice set for each individual.
- Individuals are often habitual in their behaviour and averse to new experience, which is a limitation on the trading required for the estimation of valuations of reliability and travel time.



Other issues with RP in this context have included: strong correlation between travel time and travel time reliability (de Jong et al, 2009), difficulty in modelling the full choice set and a lack of good quality RP data.

Nevertheless there have been a number of studies that have made use of RP data, the methodological aspects of which are outlined here.

Researchers in the USA have recognised the existence of high occupancy toll (HOT) lanes as an opportunity to collect and analyse RP data. The HOT lanes provide what is usually a quicker and more reliable journey time than an uncontrolled (but uncharged) alternative. This provides the basic choice situation, which is further aided by dynamically varying toll levels. Lam and Small (2001) provided an early example of such a study, where travellers were contacted after their travel had taken place to ascertain their choice of toll or non-tolled route, and the time of travel. This information was combined with supply side data to produce the choice set, which was then modelled using a variant of MV model via a mixed multinomial logit model. Although Lam and Small and similar studies (such as Liu et al, 2004) reported VORs that appeared reasonable (see Table 2.3), there were some methodological issues which remain unaddressed. There was a reliance on the assumption of choice under risk, rather than uncertainty i.e. that travellers would be fully aware of the travel time distribution associated with each of the travel options available to them. This assumption is common in the RP research but nevertheless is unrealistic and requires acknowledgement where used (as was the case in the HOT lane study of Brownstone and Small, 2005). For example, if a traveller overestimated the standard deviation of travel time on an unused route, a choice model would correspondingly underestimate their aversion to travel time risk.

There was another issue, common to Lam and Small and other HOT lane studies, that the paucity of the highway supply data was not fully acknowledged. This issue was that actual travel times were not observed, but rather flow data from loop detectors were used with standard formulae from the engineering field to convert these flows into travel times. Alternatively, research staff drove the routes and recorded travel times, but the sample sizes of this process were acknowledged as low. These HOT lane RP studies have also tended to assume that only two choice options exist, when other highway

routes or public transport options may also exist for persons making the trip. Some RP studies have however attempted to include mode choices in addition to the choice of using a toll road (e.g. Bhat and Sardesai, 2006).

Some RP reliability studies have been conducted in alternative contexts to these tolled/un-tolled examples. One such example (Prato et al, 2014) made use of GPS data from vehicles as a possible alternative to estimating mean travel times on the highway network. However practical issues remain with this approach: in many cases the sample size of observed travel times on highway links was insufficient. The method also relied upon a transport model to generate alternative routes and travel times, which risks introducing highway modelling errors into estimation of the choice model. Using travel time outputs from a transport model as inputs for an RP model has been conducted previously (Börjesson, 2008), but these were validated with travel times obtained using automatic number plate recognition (ANPR) cameras. Börjesson found her own RP estimates of the RR to be lower than combined RP/SP estimates, in agreement with the finding of Brownstone and Small (2005). Börjesson justified a preference for RP in that it represented a longer term adaptation to road conditions.

Nassir et al (2015) used farecard data from six bus OD pairs in Brisbane, Australia to estimate a choice model which included the standard deviation of travel time. Utilising an extended temporal duration of six months, they were able to identify the choice set available to each OD. They then were able to calculate performance statistics of each option available using the boarding and alighting times of the farecards. They utilised a binary logistic regression model to estimate a number of parameters, although a reliability parameter could only be estimated for the dataset that did not include transfer trips. A VOR or RR could not be reported as neither a cost nor a straightforward travel time parameter was estimated; travel time was a categorical variable related to the difference between the actual travel time and the fastest possible travel time on that route. Nevertheless, this study is instructive in the use of automatically collected payment data from public transport to estimate choice parameters – the subject of Chapter 5 of this thesis.

Wardman and Batley (2014), focussing on the aforementioned Mean-Lateness framework, highlighted a number of RP studies that have taken place within the rail

industry in the UK. A number of these remain unpublished (MVA, 2008; MVA, 2009; ARUP/Oxera, 2010). However the study by Batley et al (2011) was published and therefore provides an example of an RP study that focusses on estimating elasticities of demand with respect to reliability. This study utilised both AML and the Public Performance Measure (PPM) as indicators of reliability; the latter being the proportion of arrivals at the destination station within 5 (for short distance) or 10 (for long distance) minutes. These metrics, combined with passenger flow data, could be used to calculate the demand response to reliability levels. The study found that the demand response to reliability was lower than might have been expected in the literature. Batley et al concluded that whilst passengers dislike reliability (and might report a high willingness to pay to avoid it in an SP survey), in many situations the choice of responses available to the passenger are limited. This is a finding in favour of RP; that passenger demand would not be as affected by reliability as would be implied by willingness to pay studies based upon SP questionnaires.

The above RP studies, taken as a whole, demonstrate that the estimation of an RP VOR or ML elasticity is possible. They suggest that the limitations of RP can be overcome through innovative use of data; examples being toll road choice data and GPS data. Wardman and Batley (2014), drawing upon Batley et al (2011), suggest that RP based studies will result in a lower VOR or demand response to reliability.

### **Revealed Preference VOR Estimates**

One of the first studies to utilise HOT lane data was Ghosh (2001). Ghosh questioned road users on previous choices between a toll and non-tolled alternatives, and combined these answers with loop detector data to calculate the distribution of travel times. The central RR estimate of 1.17 was therefore made using a panel dataset. It also relied on a percentile based measure to represent reliability.

The study conducted by Lam and Small (2001) was similar to that of Ghosh insofar as it was based upon a percentile based measure of reliability in the context of HOT lanes. Lam and Small identified issues with making use of loop detector data to represent the supply side of the transport system – that data availability was often poor and only a general proxy for actual reliability conditions. They identified a central RR estimate of

0.66 for males and 1.39 for females. Small et al (2005), in another California-based HOT lane study, combined RP and SP data to estimate a RR of 0.65.

One RP study that was conducted away from the Californian HOT lanes was Bhat and Sardesi (2006), who conducted a joint SP/RP study in Austin, Texas. Their reported RR of 0.26 is low in relation to many others made in the field. One specific difference of this study to many others was that travel times were self-reported by respondents, which may have impacted upon the resulting estimate. Liu et al (2007) conducted a similar RP study on interstate routes in Minnesota and found a central RR estimate of 1.3 (although this varied by time of day). The range of Reliability Ratios could imply the importance of geographic location and the demographic characteristics of travellers in the area, although it is noted that methodological differences between these studies will have also played a part.

Börjesson's (2008) study was conducted in the Stockholm area, using traditional SP questionnaires along with RP data obtained from traffic cameras (and supplemented by transport model outputs). A RR of 0.74 was estimated. Of further relevance to the present thesis, Börjesson concluded that SP was less trustworthy than RP, based upon time horizon of each methodology. In summary, SP requires the respondent to make an instant choice of their preferred route, whereas an RP choice is made over a longer period of time based upon accumulated experience; it could therefore be considered more realistic.

A separate strand of RP research has been the use of GPS tracking of vehicles to observe travel times and choices. Carrion and Levinson (2010) developed this methodology on the Minnesota interstate HOT lanes (i.e. in a similar geographic context to Liu et al (2007)). Their central RR estimate was 0.91. The same authors tested a range of statistical measures in a later paper (Carrion and Levinson, 2013) and found the same RR of 0.91.

Prato et al (2014) was a notable GPS RP study insofar as it was based in Copenhagen and did not utilise HOT lanes but rather competing highways – the methodology could therefore be considered more widely applicable. The study estimated a RR of 1.5.

The final set of RP studies that will be covered are those focussed on the rail industry in the UK that have sought to estimate an elasticity of demand with respect to mean lateness. Recalling that corresponding SP values were presented in Table 2.1, a comparison of the two methods will be possible in the next section. The values shown in Table 2.2 are a subset of those calculated by Wardman and Batley (2014) using Equation 2.22 (although the estimates of  $\eta_{AML}$  by Batley et al (2011) were made directly). The values shown are split between short distance and long distance trips (both with London as an origin and destination). These values were chosen as they are the most similar to those presented in Table 2.1 and can therefore form the basis of a comparison. Both short-run (SR) and long-run (LR) elasticities are presented, representing both short term and long term adaptation to reliability in the form of Mean-Lateness.

**Table 2.2** - Summary of revealed AML Elasticities (from Wardman and Batley, 2014)

	London (Short Distance)			London (Long Distance)		
	Ticket Type	SR	LR	Ticket Type	SR	LR
SDG (2003)	Non Season	-0.03	-0.03	Non Season (to LDN)	-0.04	-0.04
	Season	-0.01	-0.01	Non Season (fr LDN)	-0.03	-0.03
Oxera (2005)	All	-0.06	-0.06	All	-0.28	-0.63
MVA (2008)	Non Season	-0.03	-0.05			
Arup/Oxera (2010)	Non Season	-0.05	-0.13	Reduced	-0.02	-0.14
Batley et al (2011)		-	-	Season	-0.02	-0.04
		-	-	Full	-0.01	-0.04
		-	-	Reduced	-0.01	-0.05
<b>Mean Average*</b>		-0.04	-0.07		-0.09	-0.21

\*Each study had equal weighting

Wardman and Batley identified the -0.63 elasticity associated with Oxera (2005) as an outlier and calculated average elasticities both with and without this value. This will be reflected in the comparison with SP values from Table 2.1 in the next section.

### 2.4.3 Summary of Studies

Prior to a discussion and comparison of SP and RP results, the remainder of the studies outlined to this point will be summarised within Table 2.3 below. Of particular

interest are the Reliability Ratios, which in conjunction with the elasticities presented in Tables 2.1 and 2.2 will form the basis of the comparison of SP and RP results.

**Table 2.3** - Summary table of key reliability studies that estimate a RR

<b>Study</b>	<b>RP/SP</b>	<b>Mode</b>	<b>Contribution</b>	<b>Reliability Ratio (central estimate)</b>
Small et al (1995)	SP	Car	Estimated both Scheduling and MV models, and included varying departure time and socio-demographic elements in survey.	2.3
Koskenoja (1996)	SP	Car	Thesis investigating the impact of occupation on attitude to travel time risk	0.75
Small et al (1999)	SP	Car	Estimated both Scheduling and MV models. Experimented with alternative forms of the Scheduling utility function.	2.51
Ghosh (2001)	RP	Car	Conducted a RP study of HOT lanes using panel data.	1.17
Hensher (2001)	SP	Car	Compared multinomial logit models to mixed multinomial logit models. Also treated reliability in separate parts.	0.67

Lam and Small (2001)	RP	Car	An early attempt to value reliability using RP only method.	RR of 0.66 for male, 1.39 for female.
Copley et al (2002)	SP	Car	Exploration of many SP presentational techniques including travel time histogram and varied departure times.	1.3
Bhat and Sardesai (2006)	RP/SP	Car/PT	This study was unique as it included mode choice in addition to route choice.	0.26
Small et al (2005)	RP/SP	Car	Demonstrated benefit of combining RP and SP in reliability context. Also observed significant taste variation across sample.	0.91 (RP), 0.45 (SP)
Hollander (2006)	SP	Bus	One of few SP based upon bus mode. Estimated parameters for both MV and Scheduling. Apparent mismatch between these approaches.	0.1
Asensio and Matas (2008)	SP	Car	An attempt to incorporate varying departure times into SP Scheduling/MV study.	0.98



Börjesson (2008)	RP/SP	Car	Direct comparison of RP and SP. Validated RP supply data with external data sources.	1.66 (RP only), 0.74 (combined RP/SP)
De Jong et al (2009)	SP	Freight (road)	Highlighted issues specific to freight and estimates a RR specific to this context.	1.24 (freight)
Li et al (2010)	SP	Car	Developed method for displaying SP. Models commuters and non-commuters. Found that MV and Scheduling VORs were approximately equivalent.	1.43
Tilahun and Levinson (2010)	SP	Car	Used histograms to present travel time variation to respondents. Used mode as measure of centrality.	0.89
Carrion and Levinson (2013)	RP	Car	GPS based RP similar to Carrion and Levinson (2010). Tested a range of reliability measures.	0.91
De Jong et al (2014)	SP	Freight (shipping)	Divided reliability into two components to represent freight industry.	0.1 to 0.4

Kouwenhoven et al (2014)	SP	Multi-Modal	Represented a range of modes, demographic characteristics in Dutch VOT/VOR study	0.7
Prato et al (2014)	RP	Car	Used GPS and model data to estimate VOR. Found similar RRs in peak and off-peak.	1.5
Ehreke et al (2015)	SP	Multi-Modal	Estimated RR for a range of modes and trip purposes. Only car RR quoted.	0.7

## **Discussion of RP and SP results**

The studies outlined in Table 2.3 give a RR range from 0.1 to 2.51. The average of these (unweighted) is 1.01, which would accord approximately with the values in TAG unit A1.3 (Nov 2014). It should be noted that a recent update to the value for car has reduced the RR to 0.4 (Department for Transport, 2017).

Carrion and Levinson (2012) undertook an investigation into the factors influencing the RR, conducting a meta-analysis of previous studies. They included the following explanatory variables:

- Data type (RP, SP, RP-SP)
- Scheduling / Reliability
- Region
- Reliability Measures
- Travel Time Unit (AM / PM)
- Representation of Heterogeneity in the model (observed / unobserved)
- Choice Dimension (mode, route, joint)
- Year of the Study

Carrion and Levinson specified a linear regression model, using the above explanatory variables and the RR as the dependent variable. Only one variable was statistically significant; a dummy related to whether the study took place in California. The study was not able to estimate models which explain the relationship between the RR and the methods, data and assumptions employed to estimate them.

The data presented in Table 2.3 may allow a similar linear model to be estimated. There are 21 studies in the table with a reported Reliability Ratio. The natural explanatory variables from the table are:

- year of publication.
- method of data collection – SP, RP or both.
- Mode covered by the study.

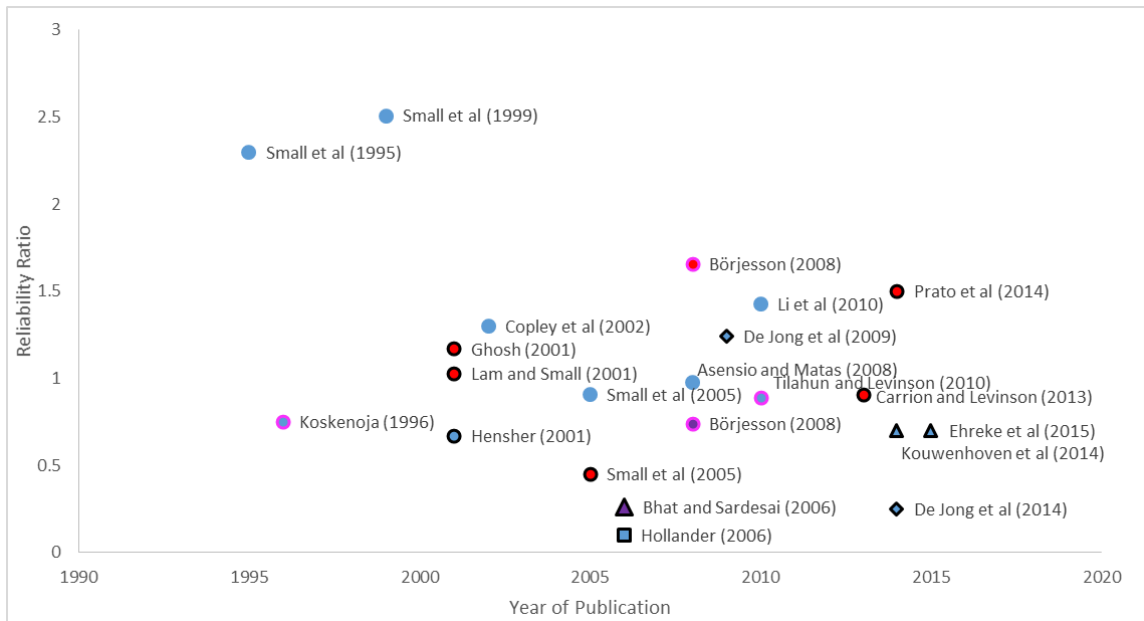
In addition, the framework used to estimate the RR can also form a categorical variable: MV, Scheduling, or both. The count of studies under each category, and average RR is reported in Table 2.4.

**Table 2.4** - Summary statistics of RR studies in Table 2.3

		Count of Studies	Average RR
Data Collection	RP	5	1.10
	SP	14	1.02
	RP/SP	2	0.73
Mode	Car	15	1.20
	Bus	1	0.10
	Freight	2	1.49
	Multi-Modal	3	1.66
Framework	MV	13	0.76
	Scheduling	3	0.95
	Both	5	1.70

Table 2.4 shows that the RP based studies in Table 2.3, on average, estimate slightly higher RRs than SP based studies. Both methods of data collection estimate much higher RRs than SP and RP combined, although the SP/RP RR is only based upon two data points. All modes appear to estimate similar RRs with the exception of bus, although this RR is only based upon a single observation. The resulting RR when both Scheduling and MV parameters are estimated appears higher than when either MV or Scheduling is employed separately, although this average is skewed by the work of Small prior to the year 2000 (Small et al 1995; Small et al, 1999).

Prior to estimating a linear regression model, Carrion and Levinson plotted the RRs against time. This is reproduced using the data from Table 2.3 in Figure 2.10.



**Figure 2.10** – Plot of studies producing a Reliability Ratio from Table 2.3

The points plotted in Figure 2.10 can be understood with reference to the lookup tables below. The shape of the data point denotes the mode being modelled, the colour is the method of data collection, and the border denotes whether the MV or Scheduling framework (or an average of both) was used to calculate the RR quoted.

Shape = Mode	
Triangle	MultiModal
Square	Bus
Diamond	Freight
Circle	Car

Colour = Data Collection	
Blue	SP
Red	RP
Purple	RP/SP

Border = Framework	
Black Border	MV
Pink Border	Sched
No Border	Both

Figure 2.10 does not indicate any clear trends. Two of the three pre-year 2000 estimates are particularly high (Small et al 1995; Small et al, 1999), which could lead to the conclusion that estimates have reduced over time. If these studies are removed however, there is no clear temporal trend. Carrion and Levinson used Ordinary Least Squares (OLS) regression in an attempt to identify the factors within a study that would influence the RR. Using the updated set of studies presented in Table 2.3 and Figure 2.10, a similar meta-analysis is conducted over the following pages. Similar to Carrion and Levinson’s study, OLS will be employed as the method for estimating the parameters of the linear model. OLS is the most common method for estimating a linear regression model and is based upon minimisation of the sum of the squared residuals.

As well as containing newer RR studies, the model estimation that follows in this chapter differs from Carrion and Levinson’s insofar as the measure of TTV, the presence of heterogeneity, and the choice dimension are all excluded as explanatory variables. Instead what remains are the four key elements:

- Year of study – a variable indicating the year of publication of a study
- Mode – a categorical variable for each mode indicating the mode of interest in the study. The “multimodal” mode variable was omitted from model estimation to avoid perfect multicollinearity.
- Data Collection – A categorical variable for each mode of data collection: RP or joint RP/SP. An SP dummy variable is not explicitly included in the estimation for the same reason as above.
- Framework – A categorical variable to represent the reliability framework used (MV, Scheduling, both), with the omission of MV.

The model estimated is therefore:

$$RR = \beta_0 + \beta_1 \cdot Year + \beta_2 \cdot RP + \beta_3 \cdot RPSP + \beta_4 \cdot Car + \beta_5 \cdot Freight + \beta_6 \cdot Sched + \beta_7 \cdot SchedMV \quad (2.28)$$

Where  $\beta_0$  is the parameter estimate of the intercept, and  $\beta_1$  to  $\beta_7$  are parameter estimates related to each of the explanatory variables. The parameter estimates of the model are presented in Table 2.5 below.

**Table 2.5** - Results of meta-analysis using OLS regression

	Coefficient Estimate	t-stat
Constant	38.452	0.917
Year	-0.019	-0.903
RP	0.616	1.554
RPSP	-0.063	-0.156
Car	-0.178	-0.367
Freight	0.168	0.377
Sched	0.439	0.983
SchedMV	1.140	2.938

Adjusted R square 0.356

Table 2.5 shows a similar result to that found by Carrion and Levinson (2012) insofar as most parameter estimates are not statistically significant at 5% (two tail test). The exception is the Sched/MV dummy variable (where separate RRs are produced using both MV and Scheduling frameworks, and mean average of the two values is calculated); this was not found by Carrion and Levinson. The positive sign on this variable suggests that studied attempting to estimate RRs using both types of reliability framework will, on average, produce higher RRs. There is not an obvious reason for this finding. If the MV and scheduling frameworks were essentially equivalent (as outlined in Section 2.3.3), it follows that no statistically significant difference between frameworks would be observed. The ‘year’ of study was negative but statistically insignificant. If the two earlier Small studies are removed and the model re-estimated, this parameter becomes positive, whilst remaining statistically insignificant. The RP indicator is positive, suggesting that RP studies will estimate higher RRs. Although its estimate has the highest t-statistic other than Sched/MV, it is not significant at 5%. The mode covered in a study does not appear to materially affect the RR produced.

Despite the largely inconclusive findings of the meta-analysis, Table 2.4 did show that there was a difference in RR estimates between those made using RP and SP. The studies utilising RP have an average RR of 1.10, compared to 1.02 when using SP data collection. This discrepancy is evidence against the hypothesis that SP studies will overestimate the VOR. It is also at odds with the conclusion of Rose and Hensher (2014) that travellers would estimate a lower VOT when they actually had experience of using the transport system in question. It should be noted however that removal of Hollander’s low RR estimate increases the average SP RR to 1.09 – close to the RP average of 1.10.

There is evidence in Tables 2.1 and 2.2 from the UK’s rail industry that RP would produce lower Reliability Ratio estimates than SP. Although the situations are not directly comparable, the RP estimates of  $\eta_{AML}$  tend to be lower. For example, the mean London – Long Distance SP estimate was -0.19. The corresponding short run RP value is -0.09, although this rises to -0.21 for long run. It should be said that the RP long run value is highly skewed by the outlying estimate of -0.63 made by Oxera (2005). When the outlier is removed, the values of short run and land run elasticities are approximately

equal. For shorter distance London travel, the mean SP elasticity is -0.12 (for non-commuters). A similar RP based value is -0.04, although again this rises to -0.07 in the long run.

### **Accounting for differences between RP and SP in the literature**

The review of SP presentation methods in Section 2.4.1 showed that a range of layouts have been used to present choice scenarios to respondents. Despite making a recommendation on the format that researchers should use, Tseng et al (2009) concluded that great care must be taken in administering SP surveys in order to prevent illogical and inconsistent responses from respondents. This suggests that SP-based Reliability Ratios have the potential to be misleading if the survey is poorly designed. This is an issue that is less relevant to RP based studies, particularly if traveler behavior is observed directly.

Börjesson (2008) reasoned that RP represented something altogether different than SP: a long term adaptation to traffic conditions which SP often did not attempt to replicate. This is supported by the RP evidence presented in Table 2.2; although the rail travellers in the studies experienced varying levels of reliability, they tended not to respond immediately. It should also be noted that if the Oxera (2005) outlier is excluded, SP-based methods will still overestimate the value of reliability even in the long run.

There are notes of caution to these findings however: as Wardman and Batley (2014) point out, travellers in the rail context may wish to modify their behaviour due to reliability concerns but not have any alternative means of travel. Therefore RP experiments where the alternatives available are poor in comparison will estimate low elasticities.

There have also been experiments that have found RP estimates to be higher than SP: examples include Ghosh (2001) and Small et al (2005). Carrion and Levinson (2012) did not find statistical significance in the difference between RP and SP RRs, but felt that the difference between RP and SP remained an open question. This is also the conclusion of the meta-analysis conducted within this chapter. A segment of the



literature has highlighted drawbacks with using SP data only, identifying a number of sources of bias. These include:

- Whether the alternatives chosen would still be chosen in reality (Börjesson, 2008)
- The range of ways the survey can be presented to the respondent (Tseng et al (2009);
- Misapprehension of the survey, particularly in the case of reliability which is a complex subject;
- Fatigue (Bradley and Daly, 1994);
- Non-trading between options (Wardman, 1988);
- Protest response (Wardman and Batley, 2014).

The present chapter also identified issues with estimating models based upon RP data: namely, the imposition of a choice set upon the traveller, and the existence of habitual behaviour (Hensher, 2010). Nevertheless, RP does overcome many of the issues associated with SP, and efforts to estimate a RP-based RR have already begun (e.g. Prato et al, 2014). In Chapter 5 of this thesis, RRs will be estimated based upon another emerging RP dataset: public transport smartcards. These datasets are introduced in the next chapter.

#### **2.4.4 Empirical Support for Equivalence**

There have been a number of studies which have made attempts to understand the empirical relationship between the MV and Scheduling EU parameters by utilising SP and RP-based methodologies. Some have been conducted prior to the completion of the theoretical work on equivalence covered in Section 2.3.3 and therefore represent a parallel attempt to establish equivalence (Lam and Small, 2001; Hollander, 2006). Others were responses to the theoretical work of Fosgerau and Karlström (2010), where their theory was tested with novel datasets (Batley and Ibáñez, 2012; Börjesson et al, 2012).

The aforementioned RP study of Lam and Small (2001) collected data on the choice between toll and un-tolled routes from travel diaries kept by commuters who participated in their study. Combined with loop-detector data, they were able to estimate an RP-based VOT and VOR using MV. Lam and Small also reported estimates of Scheduling parameters, and made some brief comparison with their estimated MV parameters. The implied VOT and VOR of the Scheduling models were lower than those of the MV model, but the authors noted difficulty in representing travel conditions for individuals beyond the section of freeway for which there was data. This unmodelled section of the highway network could have had a substantive effect upon passenger arrival times in relation to their PAT (Rose and Hensher, 2014); also this limitation would have had a greater impact on the Scheduling model than the MV specification. Furthermore, (as previously highlighted) the practice of generating a travel time distribution from flow data could be challenged. Although this was the first known comparison of empirical MV and Scheduling parameters, its issues are such that it cannot provide evidence one way or another for the theory of reliability equivalence.

The work of Hollander (2006) represented the first attempt to empirically model reliability equivalence after Bates et al (2001). The paper reported a discrepancy between the parameters of MV and Scheduling as evidence against equivalence. More generally, Hollander (2006) concluded that the MV approach will underestimate the value of travel time variation and should be rejected in favour of the Scheduling approach only. This was a strong conclusion against reliability equivalence and MV itself, however an interpretation in light of the theoretical discussion of equivalence is

relevant. The theory of Noland and Small (1995), Bates et al (2001) and Fosgerau and Karlström (2010) was dependent upon the assumption of departure time,  $D$ , being treated as a continuous variable. As the study of Hollander (2006) focused upon the bus mode, and entailed discrete values for  $D$  (and with  $D$  fixed in the SP), it is unclear that rescheduling on the part of the survey respondent was possible. Therefore the theory of reliability equivalence could not be meaningfully critiqued on the basis of Hollander (2006).

The same point can be made for the later contribution of Batley and Ibáñez (2012), which took into account the contribution of Fosgerau and Karlström (2010). Batley and Ibáñez based their analysis upon an SP survey of rail travellers, in which the hypothetical choices made were related to rail travel. It was shown that a standard deviation parameter could be estimated in the same choice model in addition to  $SDE$  and  $SDL$  parameters which would imply that equivalence did not hold. Although this was not given as a refutation of the theory of equivalence in itself, the authors suggested that equivalence would be dependent upon the form of the relevant utility functions (which will be discussed in Chapter 6). However, in addition to this conclusion, the same comments as for Hollander (2006) might be made: the public transport context (with discrete departure times) and SP surveys (when the value of  $D$  is imposed) are outside the scope of the theory of equivalence. In the case where  $D$  could be taken as continuous, the estimates of Scheduling and MV parameters appear to imply a similar VOR (e.g. in Li et al, 2010). The empirical study of Small et al (1999) also found that the standard deviation parameter was small in magnitude, statistically insignificant and not of the expected sign when Scheduling parameters were also present, therefore implying that the Scheduling parameters accounted for the entire cost of travel time variation. Recent evidence produced by Li et al (2016) proposed and found empirical support for the inclusion of a measure of variability within the Scheduling framework.

Börjesson et al (2012) based their SP study upon public transport based commuters. The motivation of the study was to overcome the known difficulty of transferring MV parameter estimates from one context to another because of the shape of the travel time distribution (Bates, 2009). Börjesson et al suggested that an improvement was to estimate Scheduling parameters and then use the theory of equivalence to transform

them into context-specific MV parameters. Again, a discrepancy was found when both sets of parameters were estimated empirically based upon the dataset (although in contrast to Hollander (2006), MV overestimated the cost of reliability compared to Scheduling). Börjesson et al (2012) went some way towards accounting for this result: arguing that when information about travel is known in advance, travellers are likely to reschedule activities to fit their actual arrival time. This acknowledgement of flexibility on the part of the travellers is crucial to understanding the result of Börjesson et al and others – the theory of equivalence relies on flexibility of the traveller in their departure time. Without an acknowledgement of flexibility in SP surveys, reliability equivalence will not be observed in practice. One such method to overcome this issue is to redesign SP surveys so that respondents indicate an ideal departure time for a given PAT and travel time distribution (using a method similar to that developed by Copley et al, 2002). Another is to utilise RP based estimates of reliability, where the departure time indicated by the data can be assumed to be optimal. In the present thesis the latter approach will be taken, based upon public transport smartcard datasets. The literature on public transport smartcards will be provided in Chapter 3.

## **2.5 Summary and Conclusions**

To conclude this chapter, an overview of its contribution with respect to the original objectives of the thesis is provided in Table 2.6.

**Table 2.6 - Objectives met by Chapter 2**

Objective	Description	Addressed in:
O4	To develop a methodology for estimating a VOR using smartcard data.	In addition to introducing the key reliability framework in detail, Section 2.4 provided a review of previous empirical studies in the field. This included an introduction to SP and RP in turn, highlighting issues with each data collection methodology. A comparison of the results from the frameworks made the case for further investigation into RP.
O5	To review the origins of the Mean-Variance framework from its origins in finance to its transition and use in transport contexts.	EUT was introduced in Section 2.1. A thorough introduction to MV, and its link to EUT and finance was provided in 2.2.
O6	To explore improvements to the standard Mean-Variance framework, including other statistical indicators of risk, the shape of the utility function and potential alternative frameworks.	Research which has examined alternative indicators of travel time risk was discussed in Section 2.3.3. This will be further developed in the next chapter. Alternative frameworks were introduced in 2.3, and the theory of equivalence between Scheduling and MV was explored in 2.3.3. The empirical

		research that has attempted to find supporting evidence for equivalence was explored in 2.4.4.
--	--	--

In addition to meeting these stated objectives, the key literature areas that have been investigated provide a background for much of what follows in this thesis. This background is summarised here.

The in-depth discussion of MV not only provides a unique background to the framework, tracing it back to its roots in finance, but will inform discussion of alternative reliability indicators. Jackson and Jucker (1982) are often credited for introducing MV into a transport context, but this literature review has shown that this introduction was primarily empirical and lacked theoretical rigour. Polak (1987a) and Senna (1994) provided the microeconomic basis for MV in transport, but it remains unclear whether the mean and variance are essential components of the approach. Markowitz (1959) suggested that alternative variables to the mean and variance may be more suited to real-world applications, but this insight has not been widely accepted in transportation contexts. Nevertheless, transport authors have begun to specify alternative measures of risk (Van Lint et al, 2008; Lam and Small, 2001).

The review of empirical evidence on the VOR highlighted some specific issues. It has been shown that drawbacks exist with the dominant SP methodology. These are both generic (e.g. cheap talk) and specific to reliability studies (e.g. high levels of protest responses). It has been shown that the alternative RP methodology is less utilised due to lack of data and pitfalls of its own, including imposed choices, difficulty specifying the choice set and habitual behaviour. Nevertheless, the literature has suggested that RP should be favoured where possible (Bates et al, 2001), and that it would be likely to yield lower RR values due to the incorporation of travellers' adaptation to actual travel conditions (Brownstone and Small, 2005). This was not borne out based on the values of RR presented in Table 2.3, although it was supported by a comparison of Mean-Lateness elasticities estimated within the rail context.

The theory of reliability equivalence was introduced in Section 2.3.3. The next step of applying reliability equivalence in empirical study was investigated in Section 2.4.3, where it was found that this has proven problematic. Whilst findings of studies such as that of Li et al (2010) appear to support the theory, Batley and Ibáñez (2012) and Hollander (2006) report findings that apparently contradict it. Lam and Small (2001)

and Börjesson et al (2012) note methodological issues in comparing the two frameworks.



## Chapter 3 - Datasets

### 3.0 Introduction

As described in the two previous chapters, this thesis will make extensive use of smartcard data provided from London's public transport network. The primary purpose of this chapter is to provide the reader with an understanding of the data that has been analysed in the course of the research undertaken here. The first part of this process will be to introduce the field and provide a literature review of key smartcard research (including research making use of Oyster card data). The second part of the chapter will introduce the datasets that will be used in the thesis. The key dataset discussed is the so-called '5% sample dataset' – this will form the basis of the majority of the empirical work in subsequent chapters. The reader will be made aware of aspects of the dataset which presented challenges for the research.

Also provided in this chapter is a description of other datasets relevant to the thesis. This includes a dataset containing 100% of Oyster records on a small section of London's public transport network. The size of this dataset potentially allows for a greater level of detail in analysis, but a downside is that it contains only one mode and a limited number of OD pairs.

In the final part of the chapter, use of the key datasets will be demonstrated; initially through some preliminary analysis. Subsequently, analysis related to different indicators of travel time reliability will be presented. The calculation of reliability is a key element of using the smartcard data in this thesis, and the empirical comparison of indicators is a theme that will be further explored in Chapter 6.

### 3.1 Background to Smartcard Ticketing

#### 3.1.1 Introduction

London's Oyster card electronic ticketing system uses technology that is more broadly referred to as a *public transport smartcard system*, of which there are numerous examples worldwide. Each system will reflect the fare policies unique to the issuing transport authority; for example Seoul's T-Money card requires bus passengers to

register both access and egress to the vehicle (Park et al, 2008), whereas London's bus passengers only register their access to a vehicle as a flat fare is levied. Such differences in fare policy have implications for the data available from the systems, which in turn will shape the research questions that might be answered from interrogation of the data. Despite these differences, the systems retain sufficient similarities such that London's Oyster card can be referred to using the generic terms *public transport smartcard* or more simply *smartcard*.

Smartcard technology has existed since the late 1960s, and has been developed and applied in many contexts completely separate from transportation: health care, banking and human resources to name but a few (Pelletier et al, 2011). The card itself is usually a similar size to a credit card; the 'smart' aspect is a microchip embedded in the card which is capable of storing data, providing read/write access, and in some circumstances processing data. Crucially, all of these tasks can be conducted securely by interfacing with the hardware of the authority responsible for the smartcard (Blythe, 1998). In the public transport context, speed of use has been a key characteristic of smartcard technology; consequently, contactless systems have become the norm (Blythe, 2004) where the cards are not required to be physically inserted into a reader.

Applications of smartcards in the public transport context began in earnest in the 1990s with successful roll-outs in Asia (Blythe, 2004). In the UK, the white paper 'A New Deal For Transport' (DETR, 1998) explicitly recognised the contribution smartcards could play in a modern public transport system. This was followed by creation of the 'Integrated Transport Smartcard Organisation' (ITSO) as a response to the concern that the de-regulated nature of local public transport services would lead to a lack of interoperability between different smartcard schemes. Whilst many smartcard schemes have been rolled out throughout the UK, it is London's Oyster card which is often cited as an example of success.

The Oyster card was launched in 2003, although this initial phase covered monthly and annual passes only. In the period since, take-up has increased due to the card covering a wide range of fares, modes (notably national rail), and geographies. It is also the fare policy of Transport for London (TfL) that Oyster card fares are lower than cash payment equivalents to encourage use of the former. This is in line with the key finding

of the thesis of Xu (2007), who forecast that smartcards would be taken up by the majority of PT users if there was a reduction in financial costs and improved convenience to the user. Other benefits include a fare capping system, which will be discussed in a later section of this chapter. Accordingly, TfL estimated that by the middle of 2012, 43 million Oyster cards had been issued, which accounted for 80% of all trips on the public transport network in London. Early research in the area has suggested that high usage levels would enable public transport authorities to utilise smartcard data to understand more about their customers (Blythe, 1998; Bagchi and White, 2005).

### **3.1.2 Literature Review**

This section of the introduction to the data will aim to give the reader an overview of the key work that has attempted to make use of smartcard data. Initially the studies will be drawn from systems worldwide, however the review will go on to focus on work related to the Oyster card in London. The limited amount of smartcard research which investigates reliability will also be discussed.

One of the first prominent examples of a successful smartcard system was the Octopus Card from Hong Kong, introduced in 1997; another was Seoul's T-Money in 2004. As such, the research making use of smartcard data tends to be recent, despite some authors identifying possible uses for public transport smartcards during or soon after their implementation (Bagchi and White, 2003; Utsonomiya et al, 2006). What is consequently found is an evolving research area where much of the work is initially focussed towards reporting supply and demand statistics of the transport systems in question – although more complex work is now emerging. A key determinant of research in this field is what data is available to the researcher, and this will be evident in the papers discussed here.

Some of the earlier studies utilised datasets which were less than ideally suited to in-depth analysis of a complex urban transport system. Nevertheless, this did not mean that valid insights into system performance and passenger behaviour could not be ascertained through research. Utilising two full days of data from the public transport system in Seoul, Park et al (2008) were able to produce aggregate statistics on numbers

of transfers, travel times, as well as demand for modes and routes. The work was validated against official figures, with the authors reporting encouraging results for the future of smartcard data in the analysis of public transport systems.

Bagchi and White (2005) utilised data from a concessionary pass scheme and a separate commercial smartcard scheme in the UK. Despite a limited dataset, an interesting insight from this work was a discussion around establishing a rate of turnover of passengers using the smartcards. Such a finding would not only be of value to public transport authorities, but could provide a useful underpinning to further research in this field. A similar concept was investigated by Morency et al (2006) using a fuller sample of smartcard data which included 277 concurrent days. In this work they reported a metric ‘new stop’ which was an aggregate measure of passenger boarding at a public transport stop for the first time – representing ‘take-up’ of a route or the system as a whole. They did not however consider a ‘drop-off’ rate which would be required to have a fuller understanding of turnover of passengers.

More recently there have been attempts to track patterns of travel within large smartcard datasets. Ma et al (2013) used a range of data mining techniques to identify trip chains. They were able to identify regular travellers and ultimately construct a door-to-door OD matrix for different user types. Tavassoli et al (2016) go one step further and compare OD matrices calculated from smartcard data with those produced by a four stage model. They found substantive discrepancies between the two data sources but do not go so far as to conclude which should be favoured. Other authors have noted that smartcard data provide little behavioural context to the trips being observed (e.g. trip purpose) and have therefore sought to understand more about travellers. Kusakabe and Asakura (2014) combined smartcard data with external datasets covering the likes of trip purpose. Lee and Hickman (2014) drew upon transaction data (often available alongside smartcard data) in combination with heuristic rules in an attempt to understand more about travellers and their trip purpose.

Whilst attempts to use the smartcard datasets have been numerous and to a large extent successful in demonstrating potential applications (see Pelletier et al, 2011 for a useful overview), authors have also been circumspect in their recommendations; often identifying flaws or drawbacks with the data. It has been suggested that situational

factors may affect the quality of data available: for example data collected from open stations or systems (where there are no access and egress gates) may be of poorer quality as passengers are less likely to use the system correctly (White et al, 2010). Another issue relates to how fares are calculated: in Singapore, bus passengers are required to use their smartcards as they access and egress from the vehicle due to all fares being calculated by distance (Chakirov and Erath, 2011). In many other contexts, such as London, flat fares are charged for bus usage and therefore egress movements are not recorded. Whilst the latter approach brings simplicity and facilitates quick vehicle egress, it does not allow easy analysis of bus passenger destinations (White et al, 2010).

### *The Oyster Card in Research*

These issues notwithstanding, there is a significant and rapidly growing body of work concentrating on London using Oyster card data as a basis for analysis, primarily carried out at the Massachusetts Institute of Technology (MIT) in the USA. Many practical considerations for using this data were described in work by Chan (2007), who compared the existing TfL methodology for OD estimation with a methodology for estimating an OD matrix using Oyster card data. The same work also identified a methodology for monitoring reliability on the underground system and discussed the excess journey time (EJT) metric in some detail; EJT being the time taken to complete a trip over the scheduled time. EJT was also the focus of the report by Frumin (2008), where this measure was derived from Oyster card data for users of the London Overground system and compared to the 'Public Performance Measure' (PPM) used on the British railways. The work went on to look at passenger awareness of public transport schedules and its impact on actual passenger behaviour. At the University of Leeds, Christos (2011) utilised a sample Oyster card dataset to report London Underground supply and demand statistics, with a particular emphasis on reliability, but did not test the influence of reliability on demand.

Similar to Chan (2007), Seaborn et al (2009) took an existing TfL manual survey methodology and contrasted it with a method of utilising smartcard data to perform the same function: in this case focussing on transfers between modes. Seaborn et al identified possible problems with linking trips within the Oyster dataset, but were nevertheless able to define transfer time thresholds for linking trips, which is

particularly relevant when data from the bus network is entry point only. Despite this research, the nature of transfers between Underground lines once a user has entered into the system using their card, but before they exit remains a region of uncertainty and a possible area for future research. Kurauchi et al (2014) found that although Oyster usage was only recorded by passengers upon entering a bus, this gave an indication of which bus route they used. After identifying travellers with repeating travel patterns (i.e. commuters), they tested whether there was variability in route choice. They found that traveller behaviour was more complex than expected; many passengers with stable origin times and locations utilised different bus services (often with differing routes) in order to reach their destination. If bus route reliability were to improve, then one should perhaps expect the variability in route choice to reduce.

### ***The Oyster Card and Reliability***

Recent research at MIT has begun to utilise Oyster card data to investigate aspects of public transport reliability. Such work has developed upon the approaches to monitoring reliability proposed by Chan (2007), for example by differentiating between types of public transport disruption such as recurrent and incident-related (Uniman, 2009). The same research has also analysed Oyster card data using a linear regression methodology to explain the causes of reliability, suggesting that travel time, interchange and incident occurrence were key explanatory variables. Whilst such a model represented a clear development in the effective use of Oyster card data, it is likely that this particular approach omitted key variables: service headway and passenger demand for example. This will be examined in the next chapter of this thesis.

Other research has begun to incorporate other data sources into Oyster card analysis (such as automatic vehicle location (AVL) data) in order to better understand public transport reliability. The most recent work in this field (Schil, 2012) endeavoured to create a unified way of reporting reliability between modes for both operational and passenger use.

### ***Using Oyster Card Data***

In utilising the Oyster card data and overcoming the associated pitfalls, the aforementioned research projects present a basis for future work using similar datasets.

Each of the London-specific studies was able to demonstrate a unique and useful methodology for using the data. However, it should be noted that a common conclusion was that even the rich and abundant dataset would still require some enrichment from more traditional surveys (Chan, 2007; Christos, 2011). In a similar vein, it is also clear that a possible improvement to the data could be the utilisation of other service quality data e.g. signalling or Automatic Vehicle Location (AVL) to complement the Oyster card records and give an understanding of exact boarding locations (Seaborn et al, 2009; Schil, 2012) or the proportion of a trip spent in vehicle (Frumin, 2008). The issues surrounding use of the Oyster data will be further investigated in Section 3.3 of this chapter.

## **3.2 Datasets**

This section will provide a brief overview of the datasets provided to the present study by TfL. The 2008 datasets were made available at beginning of the study but not all were utilised. The 2011 and Edgware to Camden datasets were provided subsequently in 2012.

### **3.2.1 TfL Datasets**

The sample datasets outlined in this section represent all trips made by 5% of the active users on London's public transport network in a given month; the data was collected during a single non-school holiday month in 2008. Subsequently an additional 5% sample dataset was provided to this study based upon data collected during a non-school holiday month in 2011. Therefore early analysis conducted as part of this study was based upon the 2008 dataset. The analysis conducted in the second half of the study was mainly based upon the 2011 data. The 2011 data were preferred as analysis suggested it contained a lower number of errors and a more reliable indication of mode choice which would prove invaluable for later work. The processes of error checking and dealing with such errors are described in subsequent sections of this chapter.

An additional dataset consisting of **all** Oyster trips (i.e. 100% of users) collected in a month on a small part of the London Underground (LU) network (the Northern Line, between Edgware and Camden Town in both directions) was also provided by TfL. This is described and some indicative analysis is conducted with the dataset. It will form the

basis of the work at the end of this chapter, where differences between statistical indicators of reliability are investigated.

### **Dataset 1 (2008) and Dataset 2 (2011): SEQUENCED\_JRNYS**

These are the key datasets to be used in the remainder of the study. It contains all trips made by 5% of the total Oyster card users within a single month in 2008 (Dataset 1) or 2011 (Dataset 2). Despite the different time periods, Datasets 1 and 2 collect the same variables (although the 5% of users in each sample will not be the same).

As part of the data collection, each Oyster card is assigned an ID number, thereby providing the potential of identifying trip chains and changing behaviour over time. Crucially for the analysis that follows, these datasets also identify the time, date and location of access and egress to the public transport system (access only on the bus mode) as well as the mode(s) used by a passenger to complete their trip. It has been shown earlier in this chapter how these aspects of the dataset have been exploited by previous researchers to provide insight into passenger behaviour on the network. Table 3.1 provides further detail on these key datasets including a brief description of each of the data fields it contains.



**Table 3.1** - Description of fields within the ‘SEQUENCES\_JRNYS’ datasets

<b>Field Name</b>	<b>Description</b>
DAYKEY	Presented in numeric format as number of days elapsed since Jan 1st 1980
PID_ENCRYPT	Anonymised Oyster card ID number
SEQUENCENO	Numerical record of the journey leg order (can be used to link journey legs into full trips)
SUBSYSTEMID	This number refers to the mode used. Some values refer to two or more modes used on one trip
STARTLOC	Records the point at which the transport system was accessed, or the fare stage in the case of bus services
ENDLOC	Records the point at which a passenger exited the transport system. This is not recorded for bus trips
ROUTE_INNERZONE	For the purposes of fare calculation, records the innermost zone of a trip
ROUTE_OUTERZONE	Records the outermost zone of a trip
ROUTE_DISTANCE	Distance between origin and destination
TB_CTEN	Start time of the trip (when the Oyster card was used to access the system) in minutes past midnight
TB_CTEX	End time of the trip (when the Oyster card was used to egress from the system) in minutes past midnight
JNYTYP	Ticket type used for the trip - either prepay, pay as you go (PAYG) or a mixture
DAILYCAPPINGFLAG	Y/N as to whether the fare for this trip was capped (i.e. maximum charge for the day had been reached)
CAPPINGSCHEME	The scheme under which the journey was capped
FULLFARE	The full fare cost of the trip
DISCOUNTEDFARE	The discounted fare (if charged)
PPTPRODUCTCODEKEY	Code relating to type of concessionary pass (if used)
PPTTIMEVALIDITYKEY	Code referring to the duration of a travel card e.g. weekly, monthly etc.

A key element of the fields named in Table 3.1 is SUBSYSTEMID, which indicates the mode or combination of modes utilised for a given trip. This represents a key choice dimension and will be useful for estimating discrete choice models in Chapter 5. The mode indicated by this field can often be determined straightforwardly where only one mode is available at a station. However, when two modes are available (thus yielding a choice), it is determined by the gate utilised at the station, as well as intermediate points on the trip where the Oyster Card is scanned to indicate the route used.

### **Dataset 3: TP\_GATESTATS**

This file contains entry and exit counts at each station by hour, day and ticket type. As such it provides a useful insight into total demand at each station. Unlike all other datasets 1 and 2, it contains counts of Oyster card and non-Oyster trips. It does not however present any information on where the entrants are going to, or where they have come from. As the dataset is focussed on stations, it consequently does not include records relating to bus based trips.

### **Dataset 4: UNDERGROUND\_SCHEDULES**

This dataset contains stops to stop distances and scheduled journey times for the entire London Underground network. These scheduled times are split by AM peak, Inter Peak and PM peak periods. This dataset is specific to 2008 operating conditions. This means that it is not an exact representation of conditions for when the later smartcard data were collected (2011 onwards). Nevertheless, comparison with the TfL journey planner in 2012 showed that it remained a close approximation of services levels, which in general do not substantively change over short periods of time.

### **Dataset 5: EDGWARE\_CAMDEN**

This dataset was provided by TfL as part of a separate data request made for the purposes of this research. It contains a record of all Oyster card trips made within a non-school holiday month in 2012 where both origin and destination of the trip lies between Edgware and Camden Town (inclusive). Unlike the 'SEQUENCED\_JRNYS' dataset, there is no unique identifier for the ticket and no information related to the ticket type, capping scheme, etc. The dataset has a simple time and location stamp for each relevant

journey made. It would therefore allow analysis of the population of Oyster card users making trips on this section of the LU network only, but does not include information related to trips that begin or end outside this geographic area.

#### **Dataset 6: EDGWARE\_CAMDEN\_GATECOUNTS**

This dataset was supplied as a supplement to Dataset 5. It comprises of all the entry and exit movements at the stations contained within the same section (Edgware to Camden Town inclusive) for the same month. Gate counts are aggregated by day and 15 minute time period. The benefit of this dataset over the previous one is that trips made with paper tickets (as opposed to Oyster card) are also recorded, allowing calculation of total access and egress numbers at a station by ticket type (Oyster/non-Oyster) and over time.

#### **Dataset 7: EDGWARE\_CAMDEN\_INCIDENTS**

This dataset is supplementary to Datasets 5 and 6. It contains date, time and duration of any incidents occurring between Edgware and Camden for the same period of time. There are also notes next to each incident describing the incident, its severity and nature.

### **3.3 Dataset Issues**

This introduction of the available datasets raises some issues that may present themselves when conducting analysis. There are also issues that are related to smartcard analysis in general, and some that are specific to the fare policies of the authority concerned (in this case, TfL). Further data issues will also become apparent as the data is handled and analysed. Within this section both types of issues are highlighted. Section 3.3.1 is a description of issues based upon background information related to the datasets and the transport network; Section 3.3.2 is an outline of issues that have been uncovered in the process of data analysis.

### **3.3.1 Pre-analysis Issues**

#### **1. Activities covered within ‘Travel Time’**

This is a key aspect of the data available at present which differentiates it from most of the VOR studies referenced in Chapter 2. When those VOR studies refer to travel time variation, their definition is the variation of time spent within the vehicle and travelling toward the destination. This definition is appropriate in many cases when the mode used is the private car. However in this thesis the definition must be expanded to cover other elements of a trip. This is because TfL’s smartcard system requires rail travellers to use the card as they enter and exit the station. The travel time (and hence travel time variation) remains a sensible term to use, but it must be redefined to cover activities other than travel in vehicle.

Travel time, recorded as the difference between access to the origin station and egress from the destination station, will therefore include the time taken to access the platform from the station entry point, waiting time on the platform, time to access the vehicle, in-vehicle time, time to egress from the vehicle, and time taken to travel between the egress platform and the station exit (where the Oyster card is registered as leaving the system). The travel time may optionally include transfer time between platforms, transfer waiting time, and miscellaneous activities (for example visiting a retail unit within a station). This range of activities may partly account for the long tail of the travel time distribution observed later in this chapter.

This definition of travel time creates two key issues. Firstly, most RRs in covered in the literature review section are calculated based upon in-vehicle time. Any RRs calculated with TfL’s smartcard data will be based upon the whole range of activities outlined above. Therefore, the RRs estimated in Chapter 5 of this thesis are not directly comparable to those in the majority of the literature. The second issue is that the randomness of total travel times will come from a number of activities, and it will be unclear the contribution of each element. The regression models in Chapter 4 are an attempt to separate out the reliability impact of different elements of total travel time.

The reader is advised to be aware of this definition of travel time in the work that follows – that VORs and Reliability Ratios calculated are based upon station-to-station conditions rather than in-vehicle time only.

## **2. Incomplete trips**

This issue is caused by travellers failing to register their smartcard at a point of access or egress at the station. The system creates an incomplete trip when it records two station access transactions by a single card when no intermediate egress transaction is made in between. The same is true when two egress movements are observed on a single card without an intermediate access movement. The existence of a (unused) dataset to record these incomplete trips would suggest that they have been removed from the primary datasets. The 2008 version of Dataset 1 was queried to test whether this was the case. It was found that of the 3,816,196 London Underground only trips (LU), 133,634 (3.5%) had a start location and time, but no end location, suggesting that the removal of incomplete trips would have to be carried out manually. Further investigation into this issue revealed that 72% of these unfinished trips were made using a pre-paid travel card, demonstrating that this type of user is most likely to bypass validation on exit from a station due to no financial penalty existing for this behaviour given the card type. More generally, this is likely to present a particular problem at stations where there are no exit barriers.

## **3. Truncation of Time Stamps**

Table 3.1 contains the fields TB\_CTEN and TB\_CTEX which are timestamp fields related to time of access and time of egress from the transport system respectively. Upon entry to the system, the time of entry in hours, minutes and seconds is recorded, for example 16:53:38. However the system truncates the time so that only the hour and minute are stored, thereby losing the number of seconds. This effectively converts the data from a near-continuous variable to a discrete one. For the example given above, the system would record 16:53, or 1013 (minutes since midnight). This has been discussed in previous research (e.g. Chan, 2007) where it was concluded that all travel time records may be inaccurate by  $\pm 59$  seconds. This is particularly a problem for short trips

where the discrepancy created by such a rounding procedure will account for a greater proportion of the total travel time.

#### **4. Bus destination points**

The existence of the bus mode within Dataset 1 and other datasets allows analysis of a road-based public transport mode alongside underground and rail. The factors affecting reliability of the bus mode is likely to be distinct from rail-based modes due to interactions with other traffic. The bus mode also has a different pricing structure to the rail based modes; on buses a flat fare is charged for any distance of travel on that vehicle whereas rail trips are charged by zone. This alternative pricing structure may be useful where a cost variable is required. However, since a flat fare is charged on buses, and the primary purpose of the Oyster card is fare collection, there is no immediately available information on the location of passenger destination points, and the time of alighting. Histograms of travel times are therefore not simple to plot. Research has made some progress in addressing this issue. The thesis of Wei (2010) showed how supplementary data points could be used to infer where the user of an Oyster card alights a bus vehicle. Furthermore, use of an automatic vehicle location (AVL) dataset which contains information related to the temporal and geographic location of a bus vehicle allows calculation of an alighting time (and hence travel time).

#### **5. Fare calculation**

There are elements of fare calculation that are specific to the public transport network in London. These are of note as they might represent a limitation in generalising the analysis in this thesis to other locations.

Fare capping is the fare policy that affects non-travelcard (i.e. pay-as-you-go) users, where the cardholder's account is debited for each individual trip made. The fare capping policy of TfL is that the total cost of travel cannot exceed a given value in the 24 hour period 04:30 to 04:29. This is represented in Dataset 1 (Table 3.1) by a binary variable on each trip record, where '1' indicates that some amount of cost less than the full fare has been charged for the trip due to fare capping.

A zonal method of calculating fares is not unusual on public transport networks (similar geography-based zonal systems are used in Paris and Berlin for example), but nevertheless it is not the only method possible. For example, one of the first public transport smartcard systems to be implemented, Seoul's T-money, relies upon distance-based calculation of fares. Other transport authorities levy flat fares based upon some time limit (e.g. Brussels). The system in Seoul is capable of accurately charging for the trip made, but is more complex than the others mentioned. TfL's zonal fare charging system omits some accuracy but allows simple calculation of fares by the passenger.

## **6. Non-smartcard based trips**

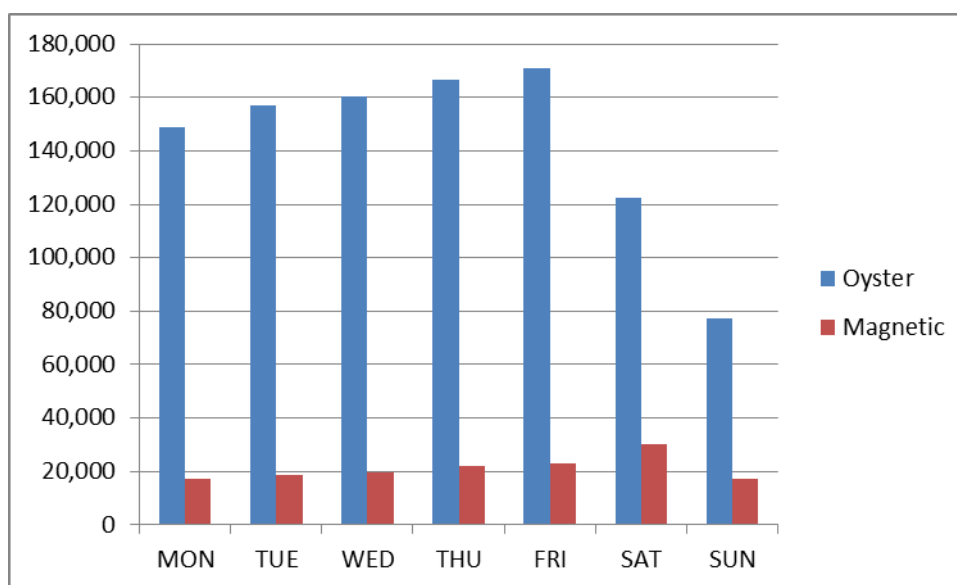
The main alternative to Oyster card/smartcard usage is referred to as 'paper' or 'magnetic' ticketing by TfL, which are trips based upon one-off cash payments and performed using a paper ticket with a magnetic strip as proof of purchase. The use of paper/magnetic ticketing is discouraged through being no longer accepted on certain modes (e.g. on buses from 2006 onwards), or charging Oyster users a lower fare compared to those using paper tickets (this measure was introduced in 2005). Such measures have resulted in Oyster card being the dominant payment method on the TfL public transport network, particularly in the time period for which this study draws its data: 2008 to 2012 inclusive. More recent developments such as contactless debit cards introduce a partial alternative to the Oyster card, but these will not be dealt with in this study. The most recent figures available suggest Oyster use accounts for approximately 80% of all trips made on the TfL public transport network. This does however mean that the Oyster card trips may not be fully representative of all trips taking place. Even Dataset 5, a population of all Oyster card trips on a section of London Underground for a given month, represents a sample of all trips made. This point is investigated using Datasets 5 and 6 (Edgware to Camden Oyster usage and gate counts respectively) and shown in Table 3.2.

**Table 3.2** - Ranked use of stations in Edgware to Camden Town study area by total number of entries and exits, split by Oyster and magnetic trips

Rank	Stations	Oyster			Magnetic			Station Total
		Entries	Exits	Total	Entries	Exits	Total	
1	Camden Town	562052	623700	1185752	172546	182516	355062	<b>1540814</b>
2	Golders Green	278264	214343	492607	16566	12584	29150	<b>521757</b>
3	Hendon Central	192690	183038	375728	13439	13470	26909	<b>402637</b>
4	Chalk Farm	186199	149516	335715	26005	20278	46283	<b>381998</b>
5	Belsize Park	184920	139095	324015	15133	11868	27001	<b>351016</b>
6	Hampstead	150165	157004	307169	10862	11803	22665	<b>329834</b>
7	Colindale	151327	120357	271684	12537	10946	23483	<b>295167</b>
8	Edgware	131313	136019	267332	10902	10335	21237	<b>288569</b>
9	Burnt Oak	136176	111807	247983	6297	5211	11508	<b>259491</b>
10	Brent Cross	66301	61755	128056	5889	5947	11836	<b>139892</b>
	<b>Total</b>	<b>2039407</b>	<b>1896634</b>	<b>3936041</b>	<b>290176</b>	<b>284958</b>	<b>575134</b>	<b>4511175</b>
	<b>Percentage</b>	<b>45.2%</b>	<b>42.0%</b>	<b>87.3%</b>	<b>6.4%</b>	<b>6.3%</b>	<b>12.7%</b>	<b>100.0%</b>

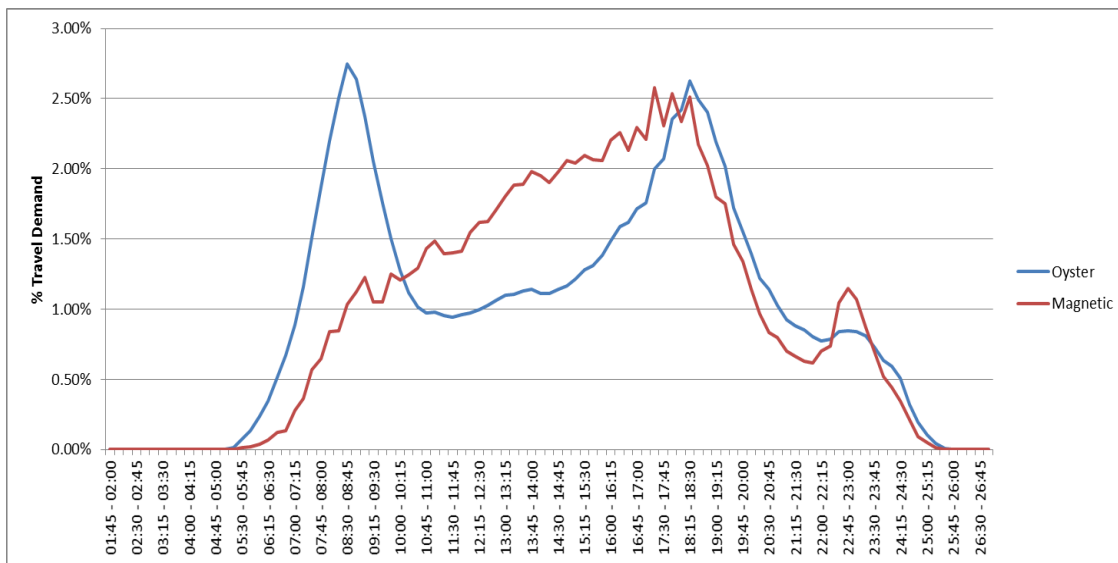


Table 3.2 shows that 87.3% of all entry or exit transactions on the Northern Line study area are Oyster based, which is higher than the generally quoted figure of 80% over the entire TfL public transport network. Of the number that are non-Oyster (i.e. magnetic) it is observed that a large and disproportionate amount (62%) come from Camden Town, suggesting it is different to the other stations in the study area and perhaps more attractive to non-regular users of London’s public transport. Figures 3.1 and 3.2 show the Oyster and non-Oyster payments by weekday and time of day.



**Figure 3.1** - Average number of entries and exits by ticket type and day (Edgware to Camden Town inclusive)

Figure 3.1 shows the dominance of Oyster usage for the area of study against other payment methods. It is interesting to note that the highest non-Oyster usage occurs on Saturday, and that proportionately non-Oyster usage is higher on both weekend days (19%) than weekdays (11%), primarily due to a significant drop in Oyster usage during weekend days. This is likely to be a reflection of a lower number of commuters travelling on the weekends – these frequent travellers are likely to use Oyster cards as their method of payment due to lower financial cost and improved convenience. It is also likely that more leisure trips occur on weekend days, with infrequent leisure travellers more likely to purchase non-Oyster tickets. The assumption that commuters utilise Oyster to a greater extent than leisure travellers is supported by the within day demand profile for Oyster and magnetic ticket types, shown in Figure 3.2.



**Figure 3.2** - Profile of weekday use by 15 minutes, normalised between Oyster and magnetic tickets

Figure 3.2 shows the difference in usage pattern between Oyster and magnetic ticket types during weekdays. It is apparent that Oyster card usage follows a recognisable peaked pattern, whereas magnetic ticket usage gradually increases through the day before declining roughly in line with Oyster card usage after the PM peak. The conclusion that would be drawn from this is that commuter and work travellers are highly likely to use an Oyster card when making their trips. The demand pattern of magnetic tickets shown in Figure 3.2, combined with their relatively high weekend use and geographic concentration of use around Camden Town station, suggest that users of non-Oyster cards are likely to be leisure travellers. This distinction between travelling groups is worthy of note in the context of this study – insights based upon the travel behaviour and travel times of Oyster users only cannot be deemed representative of the population of travellers.

### 3.3.2 Post-analysis Issues

#### Low journey times

The 2008 sample dataset contains 3,816,196 Underground-only journeys. A number of these trips, totalling 7331, had journey times which were around a minute or less. This suggests either an issue with the data collection system, or unexpected passenger

behaviour. On closer inspection it is apparent that over 96% of these trips are made from and to the same location. This could represent a passenger changing their mind about making a trip once behind the barrier, or possible instances of fraud where the passenger enters the station and then scans out without leaving in order to travel further afield for a lower fare. A small number of these low journey time trips do have a differing origin and destination station code, although these appear to take place where there are two station codes at one station – an example of which is Canary Wharf where there are DLR and Underground access points with differing station codes.

### **Sample Size**

The 2008 5% sample dataset contains 12,883,448 records – each relating to a single trip made by an individual Oyster card. Research by TfL (2011) has shown that passengers understand reliability on the routes they use but not on other parts of the network. This suggests that modelling reliability should be conducted at a level where the reliability statistics calculated are relevant to a traveller's experiences in terms of time and location i.e. at the very least on their route and at the time of day when they travel. To reflect a traveller's experience, the dataset was restricted to weekday AM peak (07:00-10:00), so as to more closely reflect homogenous conditions of travel time, reliability, crowding etc. The data was also disaggregated to the OD level, so that travel time and other supply statistics could be related to an individual traveller's trip. For simplicity, only LU trips are investigated. Incomplete trips, trips with very low journey times and duplicate trips records were identified and removed (this last point is expanded upon below). As a result of these data cleaning processes, the 2008 AM Peak LU only sample dataset consists of 667,406 records (this is a reduction from the entire 5% dataset which contained 12,883,448). This shows that limiting analysis to specific modes and temporal periods will significantly reduce the data available for analysis.

Despite this dataset containing a substantial number of records, approximately 53,000 different OD pairs were represented which meant on average there were only 12.6 records per OD pair. As has already been identified, it is often the sample standard deviation that will provide an indicator of travel time variation. To accurately calculate this value for a given OD pair, the sample size for that OD pair will need to be adequate. To illustrate this issue, the thesis of Chan (2007) suggested 20 smartcard records should

be considered a minimum to estimate a measure of the width of a travel time distribution. To model conditions that vary between days, this would require 20 records for each OD pair on each day. Only 178 OD pairs met these criteria in the 2008 sample dataset. This raises a further issue of bias when only the OD pairs with higher counts of trips are selected: these OD pairs are those that are most frequently used on the entire network. Any model which is estimated based upon these records would not be representative of the full dataset – particularly if crowding were an element of the model.

### **Duplicate Records**

Initial inspection of the data for a single OD revealed a number of records made by the same smartcard (identified by ID), with the same date and time of access and egress from the origin and destination stations. This suggested an issue with duplicate records within the wider database. A query was run in database software to identify instances with identical smartcard IDs, origin and destination points, as well as dates and times. On the full 2008 dataset of close to 13 million records, it was found that 26.9% (3,465,647) were duplicates of other records. Failure to remove such records would clearly have a substantive effect on subsequent analysis. The procedure to identify and remove duplicate records was performed on all datasets that are used subsequently (and formed part of the process detailed under the ‘Sample Size’ heading on the previous page).

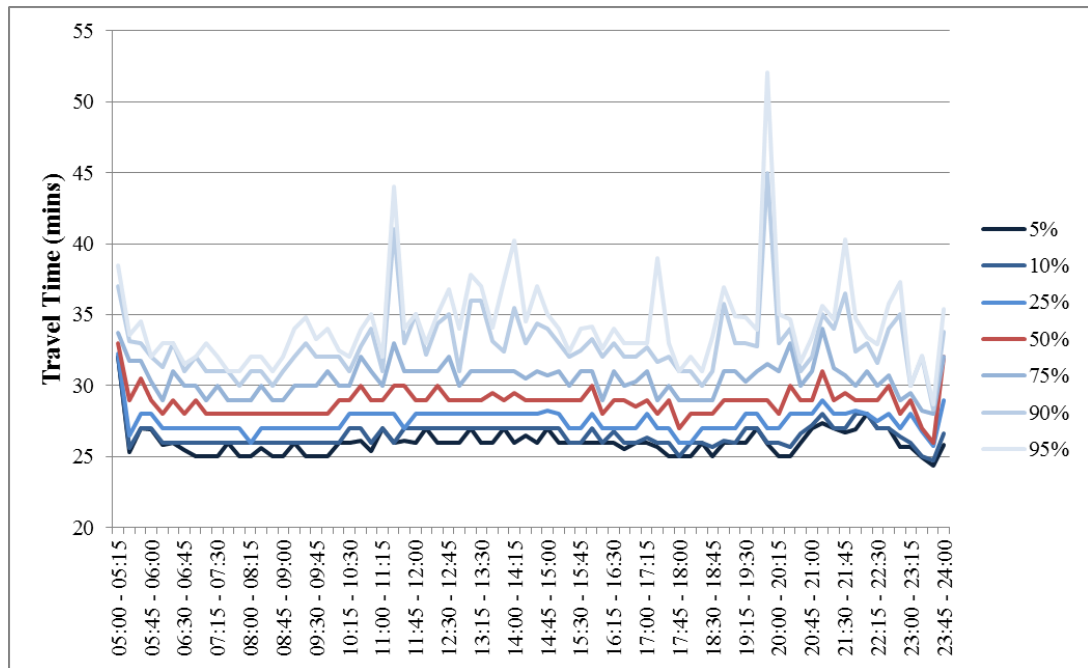
## **3.4 Initial reliability analysis**

This section will introduce some initial analysis based upon the Edgware to Camden 100% dataset (Dataset 5). The focus will be on establishing a travel time distribution and comparing common indicators of reliability used in the ‘Mean-Variance’ (MV) framework. In Chapters 4 and 5, methodologies will be developed to estimate variables and parameters of the MV model based upon smartcard data.

### **3.4.1 Travel Time Distribution**

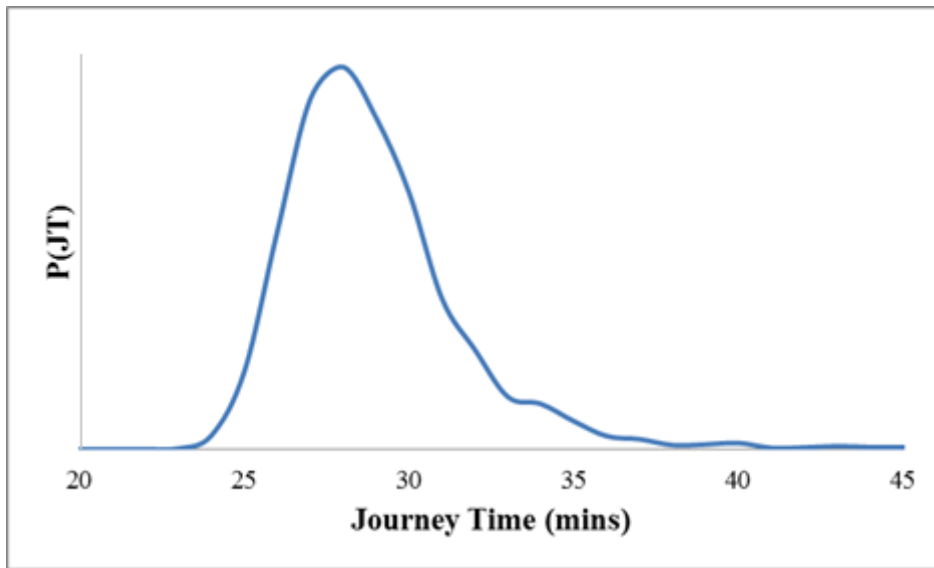
Reliability, as defined by Noland and Small (1995), is a random amount of travel time incurred by the traveller in excess to that which could be predicted in advance. This definition would therefore lead towards treating travel time as a distribution, as is

common in the literature and is formalised by the MV approach described in Chapter 2. Figure 3.3, based upon Dataset 5, gives an indication of the distribution of travel times at different times of day.



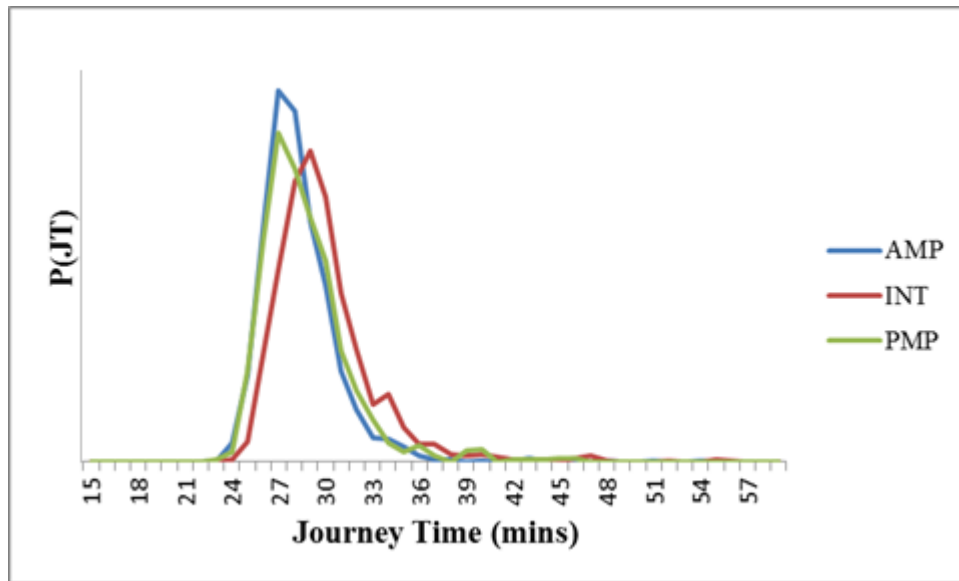
**Figure 3.3** - Percentiles of travel time from Edgware to Camden Town by 15 minute time segment

The red line in Figure 3.3 is the median travel time for each of the 15 minute time periods between 05:00 and 00:00. The travel times are longest in the very early morning and late evening time periods, and shortest in the peak periods. Of particular note are high travel times observed immediately after both the peak periods. In terms of reliability, it is the percentiles in relation to the median line that give an indication of the travel time distribution. If the distance between the 5<sup>th</sup> percentile and the median and the 95<sup>th</sup> percentile and the median were equal, this would be indicative of a symmetrical distribution. However for many time periods, it is observed that the difference between the higher percentiles and the median is greater than the distance between the lower percentiles and the median. This is indicative of a skewed distribution, with an elongated right tail. When an empirical distribution is plotted for all the weekday data, this skew is more clearly observable.



**Figure 3.4** - Empirical probability density function for all weekday travel time data Edgware to Camden Town

Figure 3.4 is indicative of a typical empirical travel time distribution that also has been found in the literature in many contexts (e.g. van Lint et al, 2008). The defining feature of such a distribution is non-symmetry and an elongated right tail – indicating that a small number of journeys will take substantially longer than the average travel time. This feature is of particular interest when dealing with travel time risk; some authors have interpreted the long tail of the distribution as the key driver of the cost of TTV (e.g. Fosgerau and Fukuda, 2010). In Figure 3.3 it is shown that the travel times in the peak periods are lower in general than those in the inter peak period. This should be reflected by the empirical PDFs plotted for each of the following periods: AM peak (07:00-09:59), inter peak (10:00-15:59) and PM peak (16:00-18:59).



**Figure 3.5** - Empirical probability density functions for AM Peak (AMP), PM Peak (PMP) and Inter peak (INT) periods, Edgware to Camden Town

Figure 3.5 confirms the result shown in Figure 3.3; the mass of the inter peak distribution is to the right of the two peak distributions, thereby indicating a longer travel time during this time period on average. The inter peak distribution has a greater proportion of mass in its right tail than the peak distributions, which may imply a greater level of travel time risk (and cost) to travellers during this time period.

Distribution fitting techniques available in standard statistical software packages do not find a statistically significant match with a pre-existing statistical distribution. Using the Anderson-Darling (AD) statistic, the closest match indicated is the log-logistic distribution. This statistic also confirms that the Normal distribution is a poor fit to the data. In the next part of this chapter a number of statistical indicators of reliability are investigated using the smartcard data to ascertain which might provide the best representation of reliability. This analysis will be further developed in Chapter 6, where these indicators will be tested using the modelling framework developed in Chapter 5.

### 3.4.2 Analysis of Reliability Indicators

In the literature review of this thesis it was shown that the primary microeconomic framework for the treatment of reliability is the MV approach. MV was developed in the field of finance (Markowitz, 1952), but has been the subject of much of the

transportation research on reliability since its first application in the field by Jackson and Jucker (1982). MV focuses on the variance (or standard deviation) of travel time as the indicator of travel time risk. In the transport literature, alternative indicators of reliability have been proposed which could replace the standard deviation or variance as the indicator of travel time risk. In this part of this chapter, a number of candidate indicators will be compared using Dataset 5. If the indicators all indicate similar levels of reliability based on the same dataset, this will provide evidence that the choice of reliability indicator is immaterial. However, divergences between what each indicator shows would imply that the choice of reliability indicator is important.

The reliability indicators used in this work will be based upon those used in the study conducted by Van Lint et al (2008), as well as those highlighted in the literature review; namely the buffer time index (of Lam and Small, 2001) and the variance of lateness, recommended for use in UK government guidance for public transport modes (TAG unit A1.3). This variance of lateness is similar to the semi-variance (SV) as defined by Markowitz (1959). All of the candidate indicators to be used are described in Table 3.3.



**Table 3.3 - Candidate indicators of reliability**

Indicator	Formula	Description
Standard Deviation	$\sqrt{\frac{1}{n-1} \sum_n (t_i - \bar{t})^2}$	Standard deviation for a given time period. This is calculated for $n$ travellers, each denoted by $i$ .
Buffer Time	$t_{95} - t_{50}$	Currently in use by TfL. Indicates the extra travel time a passenger should leave to arrive on time in 95% of cases (this level can be adjusted). It can be scaled by dividing it by the median, but loses its simple interpretation.
Late trip	$\bar{t}_{ t_i > TT_{80}}$	Calculates an average travel time of the 20% longest travel times for a given time period and/or time of day. The outcome should be compared with the mean travel time for scale.
Probabilistic	$P(t_i \geq \lambda \cdot t_{50})$	This measure is calculating the probability that travel time is above a certain threshold related to the average travel time. If $\lambda=1.2$ then this translates to “the probability that TT is greater than $t_{50} + 20\%$ ”.

Skew ( $\lambda^{skew}$ )	$\frac{t_{90} - t_{50}}{t_{50} - t_{10}}$	This measure of skew is the ratio between the distance between 90 <sup>th</sup> and 50 <sup>th</sup> percentiles and the distance between the 50 <sup>th</sup> and 10 <sup>th</sup> percentiles.
Unreliability index (UI)	$\begin{cases} \frac{t_{90} - t_{10}}{t_{50}} \ln(\lambda^{skew}) \\ \frac{t_{90} - t_{10}}{t_{50}} \end{cases}$	This measure is modified from that proposed by Van Lint et al (2008) insofar as it does not include a measure of the length of section of highway. The first of the formulae is used if $\lambda^{skew} > 1$ , otherwise the second formula is used. Both are scaled by the median travel time.
Semi-Variance	$\frac{1}{n} \sum_{t < t_i}^{t_{max}} (t_i - \bar{t})^2$	This is the variance calculated for all travellers whose travel time was in excess of the mean travel time. Taking the square root of the semi-variance results in the semi-deviation.

In Table 3.3  $t$  refers to the travel time, so that  $t_i$  is the travel time experienced by passenger  $i$ ,  $t_{50}$  refers to the 50<sup>th</sup> percentile of travel time, and  $\bar{t}$  is the mean travel time. Measures of skew and UI proposed by van Lint et al are included in the table. The semi-variance, first introduced in Chapter 2, is also added.

## Data and Analysis

These seven indicators will be calculated using a month's weekday data on the OD pair Edgware to Camden Town on the Northern Underground line in London, UK (Dataset 5). Summary statistics related to this OD pair are presented in Table 3.4.

**Table 3.4** - Weekday summary statistics for Edgware to Camden Town OD based on Oyster data only

Number of Oyster records available	5962
Proportion Oyster Usage at Edgware	92%
Intermediate stops	8
Scheduled daytime in-vehicle time	23.7
Mean journey time	29.28
Median journey time	29
Peak train frequency (/hour)	20
Inter peak train frequency (/hour)	15

The indicators of Table 3.3 were calculated for each of four time periods: All day, AM peak, inter peak and PM peak. The resulting reliability statistics are reported in Table 3.5.

**Table 3.5** - Indicators of reliability for time periods across days

Time Period	Mean	Median	St Dev	Buffer Time	Late Trip	Prob ( $\lambda=1.2$ )	Skew	UI (*100)	Semi-Deviation
ALL	29.04	28.00	3.32	6.00	4.43	6.46%	2.00	1.12	4.06
AMP	28.30	28.00	2.48	5.00	3.29	3.36%	4.00	1.92	2.51
INT	29.98	29.00	3.76	6.00	5.03	10.88%	2.50	1.69	4.45
PMP	28.77	28.00	3.20	6.00	3.41	5.19%	2.00	1.12	3.46

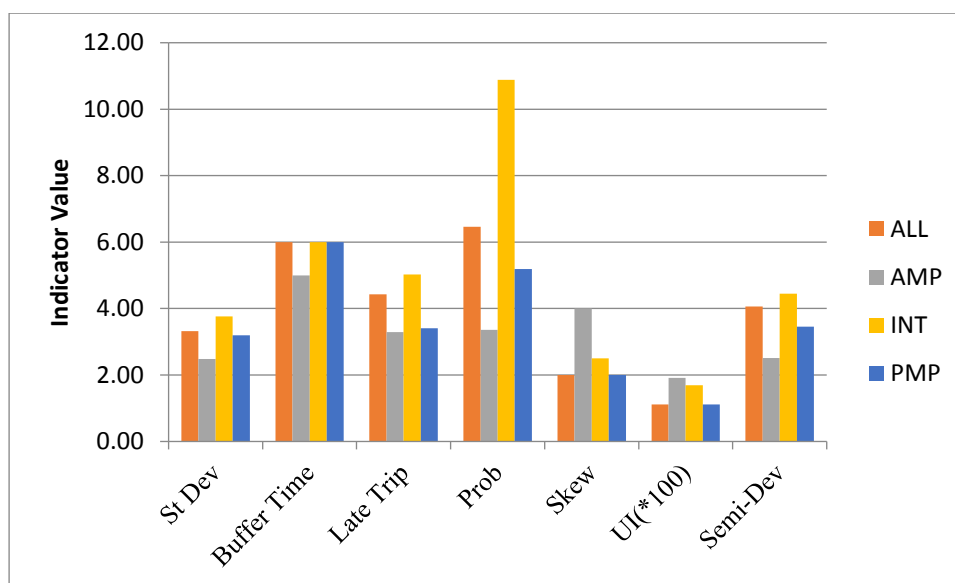
In Table 3.5, evidence of a skewed distribution exists; the ‘skew’ indicator is in excess of one in all four cases. The common existence of a skewed travel time distribution has been established both in this chapter and in the literature. The elongated right tail of such distributions has a high cost to travellers (Bogers et al, 2007). Using the sample standard deviation as the reference indicator, the AM peak period appears to have the lowest levels of travel time variation (TTV) of the three time periods. The inter peak period is the least reliable, as indicated by the standard deviation, with the PM peak intermediate to these two values.

The buffer time indicator has support in practice, being currently used by TfL as their preferred indicator of reliability. This indicator shows the AM peak to be the most reliable time period, with the buffer time indicator on inter peak and PM peak periods equal. In the light of other results, the buffer time indicator appears to be less sensitive to the shape of the travel time distribution than the other indicators.

The late trip and probabilistic indicators follow the same ranking of reliability across time periods as the standard deviation. The value of  $\lambda = 1.2$  was used for the late trip indicator as suggested by van Lint et al (2008) which implies that a trip is late if it is greater than 1.2 times the median travel time. The value of  $\lambda$  is not fixed, and it will be useful to test its sensitivity later in this chapter. The skew and UI indicators proposed by van Lint et al provide results that imply different levels of reliability to the indicators

already mentioned. The skew and UI values are highest for the AM peak, and lowest in the PM peak. This would also appear to disagree with the data plotted in Figure 3.5 where the AM peak appears to be the most reliable. On the basis of this analysis, it is difficult to support use of the UI and skew as indicators of reliability.

The final indicator, semi-deviation, approximately follows the same pattern as the standard deviation over the time periods. It is of note that the semi-deviation values are larger than the standard deviation, which is a result of representing the longer right tail of the travel time distribution. These results are displayed in Figure 3.6.



**Figure 3.6** - Chart of reliability indicator levels (ALL = all data, AMP = AM peak data, INT = inter peak data, PMP = PM peak data)

To conclude this part of the analysis, the indicators can be placed in three groups. The standard deviation, late trip, probabilistic and semi-deviation indicators all indicate the same ranking of reliability over these time periods. The buffer time indicator is not altogether dissimilar, but it is concluded that it lacks accuracy and does not differentiate between time periods. The skew and UI are the final group of indicators and present a different ranking of reliability levels that do not appear to reflect the data.

The reliability statistics are now calculated by each 15 minute time segment through the day. The level of agreement between the indicators is summarised in the form of a plot of correlation statistics of the seven standardised variables at the 15 minute time

segment level. The correlation coefficient is rank dependent and therefore adjusts for the difference in magnitude between indicators.

**Table 3.6** - Spearman correlation coefficient between all candidate indicators of reliability

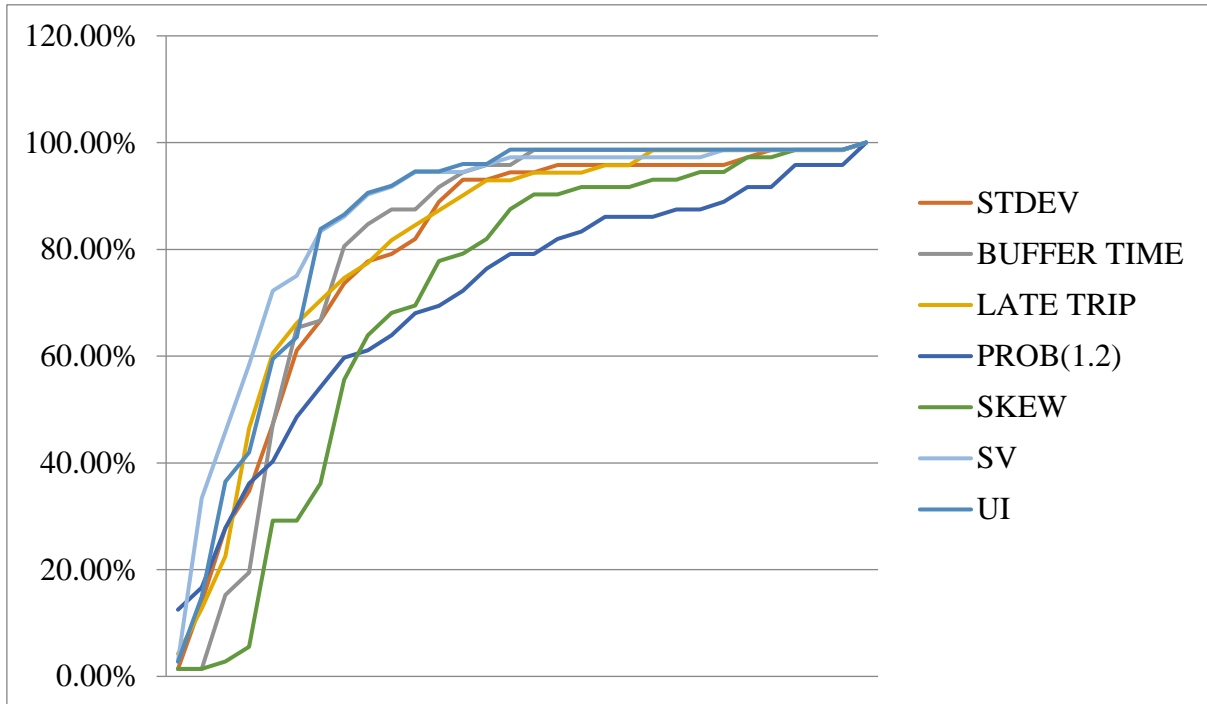
	St_Dev	Buffer_Time	Late_trip	Probabilistic	Skew	UI	Semi_Deviation
St_Dev	-	.763	.929	.765	.419	.264	.909
Buffer_Time	.763	-	.799	.857	.680	.355	.657
Late_trip	.929	.799	-	.798	.414	.281	.837
Probabilistic	.765	.857	.798	-	.702	.485	.724
Skew	.419	.680	.414	.702	-	.372	.385
UI	.264	.355	.281	.485	.372	-	.193
Semi_Deviation	.909	.657	.837	.724	.385	.193	-

All correlation coefficients are statistically significant at 5%

Table 3.6 shows a strong correlation between the late trip and standard deviation indicators. The semi-deviation also correlates highly with the standard deviation. In general there is low correlation between skew, UI and the other indicators. Of interest is that the skew and UI do not strongly correlate with one another, which is unexpected as the UI makes use of the skew indicator. The buffer time value most strongly correlates with the skew indicator, which supports the position that the buffer time indicator is useful insofar as it takes account of skew. The probabilistic indicator is the most representative of the other reliability indicators, having the highest average correlation coefficient value. However this is primarily because it correlates more highly with the UI than other indicators. This discussion is developed further by plotting the cumulative distribution of the seven indicators in Figure 3.7. This is plotted using the minimum and maximum values of each indicator over 72 15 minute time segments. It provides a greater detail on the points of agreement and disagreement between the seven candidate reliability indicators, although values on the  $x$  axis are not shown as the indicators are on different scales of travel time.

The cumulative distribution of the indicators in Figure 3.7 shows that the probabilistic and skew indicators in general do not closely match the distributions of the other indicators. Standard deviation, buffer time, late trip and SV appear to be more

closely related, but nevertheless exhibit differences which are apparent and may impact on the valuation of reliability.



**Figure 3.7** - Cumulative distribution of indicators used for within day analysis

Figure 3.7 shows that the SV indicator would have a probability density function with a greater proportion of its mass to the left of the other indicators. This is due to 22 of 72 SV values having low values: i.e. this indicator is more likely to imply a reliable trip than the others shown. There is evidence that the probabilistic indicator will imply a poorer level of reliability for the same data. However it has been acknowledged previously that the value of  $\lambda$ , which forms the basis of defining unreliability for this indicator, has been defined arbitrarily (a value of 1.2 from the literature has been used). The sensitivity of  $\lambda$  as part of the probabilistic indicators is now tested in greater detail. This is followed by a discussion on the buffer time indicator, the standard deviation and finally the semi-deviation as a viable alternative.

**Discussion of Reliability Indicator Analysis**

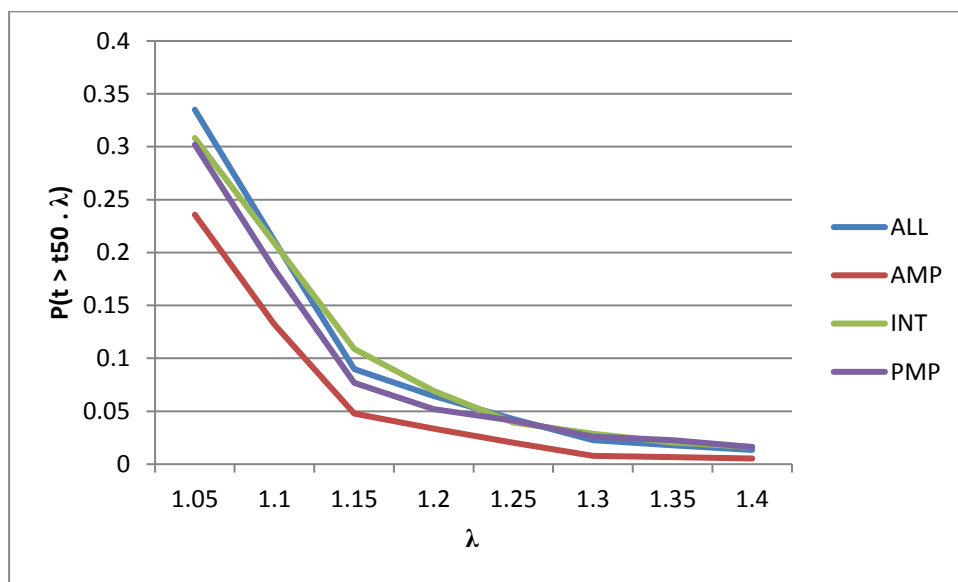
In this section, aspects of the preceding analysis will be discussed in greater detail including;



- the role and value of  $\lambda$  in the probabilistic indicator
- the practical benefits and dis-benefits of the candidate reliability indicators

The probabilistic indicator operates by using a parameter to define the threshold at which a trip is deemed to be unreliable;  $\lambda = 1.2$  has been used which was interpreted as any trip taking 20% longer than the median could be categorised as unreliable.

Returning to the time periods (as displayed in Table 3.5), the value of  $\lambda$  was varied for each of these to between 1.05 and 1.4 inclusive to show the sensitivity of the indicator. These values are plotted for the four time periods and for varying levels of  $\lambda$  in Figure 3.8.



**Figure 3.8** - Probabilistic reliability indicator for each time period, varying by  $\lambda$  parameter

Figure 3.8 shows that the relationship between the probabilistic indicator and the parameter value is non-linear. Based on all of the weekday Oyster data used in the previous section, the relationship between the probabilistic indicator and  $\lambda$  value is exponentially decreasing. That is to say, the result of probabilistic indicator is highly dependent upon the choice of  $\lambda$ , such that that differing levels of reliability will be implied based on its specification.

Another issue arising from Figure 3.8 is the difference in ordering of the indicators when  $\lambda$  is varied. The other six candidate reliability indicators examined would suggest

that the most reliable time period is the AM peak, followed by the PM peak, all day, and finally inter peak. However this does not hold for all values of  $\lambda$ ; in fact this ordering holds only for  $\lambda=1.15$  and  $\lambda=1.20$ , further suggesting that this indicator is liable to producing inconsistent results.

Buffer time indicators are a key alternative to standard deviation as a measure of reliability (Van Lint et al, 2008). It is used in practice as it can be communicated effectively to the general public; for example,  $(t_{95}-t_{50})$  calculates how much extra time above the median that a traveller should allow in order to arrive on time with 95% confidence. However Figure 3.6, based on the four time periods, suggested that this indicator was inconsistent with the others used, as it showed the same levels of reliability for three of these time periods. Figure 3.7 went on to show that this was not the case when calculated by 15 minute time segments; the range of the buffer time indicator was actually the highest of all the indicators.

Whilst being a useful indicator for passengers, the analysis above questions whether the buffer time is a useful tool in the planning and appraisal of public transport systems due to the suggested inaccuracy of the measure. In choosing just the median and one other percentile of the travel time distribution, there is the possibility that key information is missed; this indicator only takes into account lateness, but according to the 'Scheduling' framework (Small, 1982; Noland and Small, 1995) there is also a cost associated with early arrival. This criticism is also valid for the late trip and probabilistic indicators which also do not reflect earliness.

Moving on to the standard deviation measure, it is important to recognise its appealing statistical properties that have resulted in its common use in transport reliability contexts. Firstly, it is a widely understood indicator of spread and appears to be a natural choice for describing the width of a distribution. Secondly, Figure 3.7 shows that is closely aligned to some of the other indicators, which gives confidence that the estimates of reliability it produces are representative of the actual situation. Finally, as was shown in Chapter 2, it is the primary indicator of risk used in MV analysis.

Drawbacks of the standard deviation indicator do exist. Firstly, the data analysis in this chapter has shown that the distribution of travel times is skewed, and therefore the conclusions one can draw from the standard deviation measure could be misleading. Second, the measure has also been treated with caution due to difficulty in conveying it to the general public (Uniman, 2009).

The final reliability indicator of note is the semi-deviation. The semi-deviation has been introduced as a viable alternative to the standard deviation. The measure has implied similar levels of reliability to the standard deviation measure as demonstrated by a high level of correlation between the two statistics. Similar to the buffer time indicator however, the measure as applied here was focussed upon the right tail of the travel time distribution and therefore does not account for risk in earliness.

### **Implications of Empirical Work**

The main body of the work conducted in this chapter to this point has been focussed on a single OD pair on the Northern Line of the London Underground network. It concentrated on reliability measures calculated for peak periods and 15 minute time segments across all weekdays during a single month. The analysis of the seven candidate indicators of reliability can be distilled into the following three general points:

1. Different indicators have given different answers to the level of reliability during a specific time period – be that whole time periods (e.g. AM Peak) or 15 minute time segments. Although there was some correlation between indicators, the level of agreement between two indicators was never strong enough to consider one to be a close representation of another. These levels of disagreement between indicators mean that results of prediction or appraisal work on reliability will be dependent to some extent upon the choice of indicator.
2. It was difficult to objectively identify a preferred indicator that best encapsulates the reliability performance in the case of the London Underground. This is a finding similar to that of van Lint et al (2008). In particular, skewness did not appear to capture reliability levels in the same way that other indicators did and should be discounted from further analysis. The

probabilistic indicator tended towards extreme values (both high and low) and sensitivity testing of the parameter suggested that the indicator is unreliable in itself. Buffer time, late trip, standard deviation and semi-deviation were found to be more similar to one another, but it was difficult to suggest that one should be preferred over another.

3. The indicators have tended to be focussed towards representing the spread of the travel time distribution, which is effectively in-keeping with MV (trading off between travel time and travel time risk). There was some emphasis on representing the lateness of trips, but little or no recognition of earliness and how it might be valued differently to lateness (i.e. in line with the Scheduling framework of reliability).

As a result of this analysis, the Buffer time, late trip, standard deviation and semi-deviation indicators will be taken forward for further analysis in Chapter 6.

### **3.5 Conclusion**

The purpose of this chapter was to introduce smartcards and the data available to this thesis which will be further utilised in the coming chapters. This has been achieved through a combination of a literature review, description of the data and analysis of reliability indicators using the data. The chapter also contributes toward a number of objectives, shown in Table 3.7.

**Table 3.7 - Objectives met by Chapter 3**

Objective	Description	Addressed in:
O1	To apply smartcard datasets to a number of real world situations – drawing and developing upon existing studies.	The background and literature review of the use of smartcard data was provided in Section 3.1. The datasets were introduced in greater detail in 3.2. Analysis using the data was conducted in Section 3.4.
O2	To provide a critique of the smartcard data available (and smartcard data more broadly).	Based on the literature review, general issues related to the data were discussed in Section 3.3.1. Further issues coming to light after analysis were outlined in Section 3.3.2.
O3	To develop the means for improving understanding of the factors affecting Transport Reliability using smartcard data.	This chapter demonstrated how the data could be handled and analysed in order to calculate key reliability indicators. This forms the basis of the analysis taking place in Chapter 4.
O4	To develop a methodology for estimating a VOR using smartcard data.	This chapter demonstrated how the data could be handled and analysed in order to calculate key reliability indicators. This also forms the basis of the analysis taking place in Chapter 5.

O6	To explore improvements to the standard MV framework, including other statistical indicators of risk, the shape of the utility function and potential alternative frameworks.	3.4.2 formed a substantive section of the chapter – initially identifying a number of statistical indicators from the literature. These indicators were compared using the smartcard data and a subset of these will be taken forward for further analysis in Chapter 6.
----	---	--

The chapter began with an introduction to smartcard technology, which developed into a literature review on the subject. The datasets available to the thesis were then introduced. TfL provided three key datasets to this project: 5% sample (2008), 5% sample (2011) and Edgware to Camden 100%. Each of the individual datasets was introduced sequentially, noting that Datasets 1 and 2 (SEQUENCED\_JOURNEYS in 2008 and 2011 respectively) will provide the basis of much of the analysis in following chapters. The form of the datasets raised some issues for research which were outlined in Section 3.3.1. Subsequently the datasets were interrogated and further issues were highlighted in Section 3.3.2.

Analysis of the smartcard data began in Section 3.4 with a single OD pair from the 100% Oyster dataset. This dataset allowed a temporal disaggregation to 15 minute time segments and it was therefore possible to calculate a range of percentiles to represent the changing travel time distribution across an average weekday. This analysis was further developed by an attempt to fit a distribution to the data for each of the three primary time periods (AM peak, inter peak and PM peak). A substantive part of this chapter was to introduce a number of candidate reliability indicators from the literature which could potentially be used to replace standard deviation as an indicator of risk in MV. The data showed that these indicators demonstrated different levels of reliability for the same dataset, and it was not clear which should be favoured. Buffer time, late trip, standard deviation and semi-deviation all appeared to be potential reliability indicators. These will be tested in Chapter 6 with the modelling approach developed in Chapter 5.

This chapter has described the range of data available, identified some of the pitfalls that come with the use of such data, and conducted some analysis of reliability indicators which will feed into the modelling work that follows. Chapter 4 will focus on the estimation of models which identify the factors affecting reliability and the prediction of reliability levels. Chapter 5 will estimate discrete choice models based on the smartcard data, and in Chapter 6 indicators of reliability will be re-introduced and tested against actual passenger behaviour.

## **Chapter 4 - The Factors Affecting Reliability on the London Underground**

### **4.1 Introduction**

Chapter 2 of this thesis introduced the Mean-Variance framework for transport reliability in some detail. Chapter 3 introduced the data available to this thesis – namely smartcard data provided by Transport for London. The present chapter draws upon this preparatory work by attempting to estimate the level of travel time reliability on the London Underground based on the smartcard data alone. This process will identify which factors are driving these values and to what extent they predict future levels of reliability. This work comes in advance of Chapter 5, where the Mean-Variance preference parameters are estimated based upon smartcard data alone. The present chapter should therefore be viewed as the first of two parts; focussing on the supply of public transport in advance of Chapter 5 which is focussed on the demand for public transport.

### **4.2 Previous Work**

There has been a substantial amount of work conducted in this area, which can be broadly split into two parts. The first is primarily focussed upon the private car and seeks to utilise readily available variables, such as mean travel time, in order to forecast reliability. Such studies will be outlined in Section 4.2.1. The second body of work in this area is related to public transport modes; primarily bus. The public transport studies have tended to undertake bespoke data collection in order to forecast levels of reliability. These studies will be explored in Section 4.2.2. Whilst covering both private and public transport, the focus of this thesis is mainly on the latter in order to inform model estimation in Section 4.5.

#### **4.2.1 Forecasting Reliability for Highway Modes**

In a recent review in this area, de Jong and Bliemer (2015) focussed on national guidance for establishing the relationship between reliability and other variables, which will be outlined here.



In the UK, Arup (2003) established a relationship between congestion, distance of a trip and reliability. The coefficient of variation (CV), the ratio of the standard deviation of travel time to mean travel time, was estimated by the formula:

$$CV = 0.148 CI^{0.781} D^{-0.285} \quad (4.1)$$

Where  $CI$  is the ratio of mean travel time to free flow travel time, and  $D$  is the distance of the trip. This approach, used in the UK's official guidance, therefore assumes that reliability levels are explained by distance and congestion only. It is somewhat similar to that used in Australia and New Zealand, where standard deviation can be calculated using the standard deviation under congested and uncongested conditions, as well as the volume to capacity ratio.

De Jong and Bliemer's review indicated that a substantial amount of evidence has been generated from research undertaken in the Netherlands. In a straightforward example, the literature review of Besseling et al (2004) suggested factoring travel time benefits by 1.25 in order to take account of reliability. A more detailed approach was taken by Peer et al (2012), who utilised speed and demand data from Dutch highway management systems. Their study specified weather, day of the week and traffic conditions on the rest of the network as explanatory variables for reliability. A simple linear regression was utilised, which is similar to many of the public transport studies outlined in the next section. Results from these models were found to be mixed and often counterintuitive, leading the study to further investigate the relationship between standard deviation and travel time/delay. The result of this latter part of the research was instructive: the relationship between the standard deviation of travel time and travel time was positive but non-linear, with the authors concluding that fluctuations in demand may result in different relationships between speed and reliability.

In contrast, Hellinga (2011) was able to estimate a strong linear relationship between travel time and standard deviation, also using data from the Dutch highway system. It should be noted that to achieve this result non-recurrent conditions were removed from the data, which may have themselves been explanatory variables for reliability. This study is however supported by research elsewhere: Mahmassani (2011), using highway data from the USA, estimated a relatively strong relationship between

speed and the standard deviation of travel time, although also noted that some representation of non-recurrent congestion would be required.

Other work in the highway context has tended to support the non-linear relationship suggested by Peer: for example, Kouwenhoven et al (2014) estimated a reliability model using similar motorway data but specified a logarithmic term for delay. This latter study was in agreement with Arup (2003) insofar as delay and distance were the key explanatory variables for reliability, and the relationship was non-linear. Non-linearity was further supported by research that has recommended a power-law function to represent the relationship between reliability and the mean-delay (Geisterfeldt et al, 2014). Kouwenhoven et al (2005) attempted to incorporate additional explanatory variables into their modelling related to the occurrence of incidents or maintenance work. They found that these variables had no discernible impact on the fit of the reliability model. Instead they found that the relationship between travel time and reliability held even when an incident occurred.

#### **4.2.2 Forecasting Reliability for Public Transport**

The first researchers to investigate the factors driving reliability levels of a public transport system from observed data were Sterman and Shoefer (1976), who collected data on bus service performance in the Chicago area. They specified the dependent variable as the inverse of the standard deviation of travel time. Using a linear regression model, they were able to show that increases in length of bus route, extent of traffic signal control, traffic volumes and passenger boarding numbers all degraded reliability levels.

Abkowitz and Engelstein (1982) were focussed upon mean travel time in addition to the standard deviation of travel time. Like Sterman and Shoefer, they collected bus data, and were notable as they were the first to conduct work in this area using automatic data collection methods. Their model specification included two types of explanatory variable: static and dynamic. Static variables were unchanging and largely based on an inventory of the route; these were link length, count of traffic signals, length of parking restrictions and count of unsignalised junctions. Dynamic variables were those that varied between bus vehicles and included number of boardings and alightings on a section, number of stops made, time of travel and direction of travel. Predictors of travel

time were found to be the length of a section, counts of boardings and alightings (separately), the length of on-street parking and the count of signalised intersections. An adjusted  $R^2$  of 0.92 implied that this model would be a good estimator of mean travel time levels. The number of stops was omitted from the model as it had the incorrect sign and was suspected of correlating with other variables. It is likely that substantial correlation between distance, number of stops and number of boarding movements will impact upon any similar model being estimated. The model with standard deviation as its dependent variable provided a less definitive demonstration of the method; only standard deviation on the previous length of route was a statistically significant predictor of reliability on a route (route length was estimated but not significant at 5%). Again, a high adjusted  $R^2$  (of 0.89) implied a high level of predictive power of the model. The main finding was that reliability levels are correlated along a route; but this is not necessarily useful in determining the root cause.

Strathman and Hopper (1993) deviated from previous studies in that they used a discrete choice model as opposed to linear regression to model bus operations. The dependent variable was probability of on-time arrival, predicted by number of boardings and alightings, route length, location of a timing point in relation to the rest of the route, number of stops, time period, weekday/non-weekday, experience of driver and scheduled headway. The modelling found that the count of alighting movements, location of the timing point and headway were the key determinants of whether a bus would be punctual or not.

The thesis of Cham (2006) is interesting insofar as it was among the first to make use of large automatic payment collection (APC) systems to examine the factors driving reliability levels. The thesis detailed how bus location data and automatic payment collection data were collected and combined on a Bus Rapid Transit service in Boston, USA. The thesis did not develop a regression model, but rather examined the relationship between reliability and explanatory variables in turn. In addition to the number of passenger boarding movements, it was found that average deviation from schedule was a key explanatory variable.

Mazloumi et al (2008) deviated from much of the literature in this area by employing a percentile indicator of reliability similar to that proposed by Lam and

Small (2001). This study was among the first to utilise automatically collected bus location data, using GPS to estimate a linear regression model capable of predicting reliability. This study found that the length of a section would have a substantive impact on reliability. In addition, 'land use' (a categorical variable for residential/industrial) was found to be statistically significant. The same authors conducted a similar study (Mazloumi et al, 2010), where they found that the number of stops, count of signals and deviation from schedule were also factors driving the level of reliability. They also found that a model that utilised the standard deviation of travel time as a dependent variable had greater explanatory power than one that used a percentile based indicator. A drawback of the 2010 study was that it was unable to match bus location data to automatic passenger count data for greater insight on boarding and alighting movements.

Sorratini et al (2008) departed from previous studies in the methodology utilised to understand the factors impacting upon reliability. A micro-simulation software package was employed to assess reliability on a section of bus route in the city of York, UK. Through using this methodology, some new insights were achieved which would have been difficult to represent using a linear regression model only. For example, modifying the layout of busy stops had the potential to improve bus reliability. Furthermore, it was found that bus only lanes benefited reliability levels. As would seem intuitive, they also found that speeding up passenger boarding time would improve the ability of a bus service to maintain uniform headways.

Chen et al (2009) collected data over a three day period using 396 surveyors on 30 of Beijing's bus routes. They obtained over 70,000 observations at a bus stop level and utilised these data in a range of ways. They identified that route length was a key determining factor for reliability, and furthermore that reliability levels would be particularly poor when a route was over 30km in length. In common with the finding of Sorratini (2008), it was found that bus priority lanes improved reliability levels. The study concluded that although based on a large dataset, automatically collected data (vehicle location and payment data) would allow further development of the methodology.

The study of El Geneidy et al (2010) combined such datasets and used a linear regression model to estimate the level of reliability and the factors driving it. The key innovation was to link 150,000 stop level observations of buses to automatic payment collection counts. In terms of the variables used, the study defined a longlist based on those already covered in this literature review. An additional variable used in this particular study was whether a lift was used to access a boarding point. It was found that the variables impacting on travel time deviation were the length of a section, the number of stops, count of boardings/alightings as well as driver experience.

This literature review has shown that there is a substantive amount of evidence on the factors affecting reliability of bus-based public transport. The key explanatory variables affecting reliability levels are:

- The length of the section
- The number of passengers boarding and alighting
- Experience of the driver
- The existence of a bus lane (or more broadly priority)
- The difference between actual and scheduled operation
- The number of stops

The literature review also shows a developing field. Although a number of techniques are used to establish the relationship between variables, it is linear regression which is most often utilised. The literature review has also demonstrated a growing use of automatically collected data to derive these relationships. Although the studies cited to this point are based around bus operation, the results will be generalised to inform development of a linear model for the London Underground. An attempt to estimate such a model was attempted in the thesis of Uniman (2009), discussed in the next section.

### **The study of Uniman (2009)**

The work of Uniman (2009) may assist in providing an improved understanding of the factors affecting reliability in the context of the London Underground. The data available to this study was a 5% sample of Oyster card users in two different months (February and November). The study looked at reliability performance across a large

number of OD pairs and the data used were similar to the datasets available to the present thesis.

Uniman’s study hypothesised that the factors impacting on reliability (defined as the Buffer Time; the difference between the 95<sup>th</sup> percentile and median travel time, in minutes) would be:

- Journey length (measured as median travel time)
- The existence of a transfer (a dummy variable, taking a value of “1” if a transfer was required)
- The existence of an incident (e.g. vehicle breakdown) (a dummy variable, where “1” indicated that an incident occurred)
- A dummy variable to account for a seasonality impact

The regression equation was therefore proposed as follows

$$\begin{aligned}
 \text{Reliability}_{i,j,m} & \qquad \qquad \qquad (4.2) \\
 & = \beta_0 + \beta_1 \text{Journey Length}_{i,j,m} + \beta_2 \text{Transfer}_{i,j} \\
 & \quad + \beta_3 \text{Incident}_{i,j,m} + \beta_4 \text{Seasonality}_m
 \end{aligned}$$

Where  $i$  and  $j$  were defined as the origin and destination respectively and  $m$  is the day and time period.

The parameters were estimated based on Ordinary Least Squares (OLS) regression, with the coefficient estimates in units of minutes. This process found that all explanatory variables influence reliability levels with the exception of seasonality. A summary of the model estimates is reproduced in Table 4.1.

**Table 4.1** - Model estimation results from Uniman (2009)

Variable	Coefficient Estimate	T-statistic
Intercept	3.068	44.65
Journey Length	0.120	43.08
Incident (Dummy)	9.912	157.02
Transfer (Dummy)	0.978	15.39
Seasonality (Dummy)	0.082	1.69

Adjusted  $R^2=0.466$

This model therefore shows that it is median travel time, incidents and transfers that will have an impact upon levels of reliability. The study established that there is a strong relationship between journey time and reliability levels on the London Underground, with a longer journey time leading to poorer reliability levels. The existence of a transfer would also decrease travel time reliability: a single transfer increasing the difference between the 95<sup>th</sup> percentile of travel time and the median travel time by approximately 56 seconds on average, all else equal.

The incident coefficient was highly significant, which the study interpreted as showing that the occurrence of an incident would reduce reliability. This result is supported by intuition, but the methodology for determining when an incident took place was based upon the Oyster data only. When long travel times were observed during a time period, these would be marked as incident related. Another explanation might be that the service was simply experiencing a prolonged period of poor performance or unreliability. Without external verification that an incident was taking place it is difficult to conclude that the model result shown above provides evidence that an incident leads to poorer reliability.

Finally, the seasonality dummy variable was not significant at 5%, but this is unsurprising given that the two months of data used were November and February – likely to be similar in terms of weather and other general conditions in the UK. If the seasonality variable were instead comparing November and June for instance, perhaps some significant difference would have been detected.

The methodology and model developed by Uniman (2009) therefore provides a useful framework for discovering the factors impacting upon reliability on the London Underground. Using a percentile based indicator for reliability, the study showed that increasing journey time length and the existence of a transfer would negatively impact upon reliability.

The literature review of the previous section shows that the factors affecting the reliability of buses are more varied than the two key factors identified by Uniman. This

would suggest that there is value in collecting and analysing additional variables, with the objective of overcoming omitted variable bias in the regression model whilst also allowing the identification of further factors affecting reliability.

### 4.3 Model Variables

The studies outlined in this chapter so far primarily utilised a simple linear regression form such as that shown in Equation 4.2. Such model specifications have the benefit of allowing the parameters for a number of explanatory variables to be estimated simultaneously. The parameter estimates in a standard linear regression model are interpreted as the impact that a change in an explanatory variable has on the dependent variable, holding all else equal. Ordinary Least Squares model estimation will be utilised in line with the models estimated in the literature review.

The purpose of this section is to give a fuller account of the factors impacting on reliability levels than has been presented to date and develop a model capable of forecasting reliability levels. This will involve testing additional explanatory variables with the general model specification in Equation 4.2. Building upon initial investigation of the data from the previous chapter, the following variables are calculated:

**Reliability** – Chapter 3 showed that this can be defined and calculated in a number of ways. The primary indicator for reliability used here will be the standard deviation for a given OD pair. This is, the standard deviation of all travellers between a single origin and destination, across weekdays and for a single time period. It should be noted that all journey time records are from the origin gateline to the destination gateline i.e. incorporating a range of activities outside in-vehicle time as described in Chapter 3.

**Mean Journey Time** – Uniman (2009) showed that the median journey time was a key explanatory variable for reliability. As this thesis is primarily focussed on the Mean-Variance model, the mean travel time will be used. This will be calculated in a similar fashion to the standard deviation; the mean travel time of all trips observed through the data on a single OD pair across weekdays and for a given time period.

**Distance** – The distance travelled will likely be a strong predictor of reliability. Stop-to-stop data (measured in km) has been provided by TfL for the purposes of this



project (described in Chapter 3). This allows track distance to be calculated for any OD pair. An issue arising is the likelihood that mean journey time and distance will be strongly correlated, which will present issues for estimating parameters on these variables. A possible solution may be to specify only one of these variables in model estimation.

**Headway** – For the purposes of this process, headway will be defined as the average scheduled temporal spacing (in minutes) between vehicles on a line during a given time period. It can be calculated using the TfL journey planner and published timetables. Where a transfer occurs on a trip, the headway of the relevant service at the two departure stations will be added together.

**Demand at Origin Station** – This variable is a proxy for the number of people using a station. It is based upon the gate count dataset described in Chapter 3 and is calculated by taking an average daily gate count (in both directions) at a station for a given time period. This value is then matched based upon the origin station of a trip. The purpose of this variable is to represent the possible crowding effects for a passenger as they attempt to move from the gateline to their departure platform and then onto their train.

**Demand at Destination Station** – This variable is calculated in the same way as ‘Demand at Origin Station’, but is assigned to destination stations. It is used to reflect potential increased variability in travel times between the platform and the exit gateline due to high volumes of other travellers.

**Count of Intermediate Stops** – This variable is a count of the stops between the origin and destination. In the bus context it has been found to be a good predictor of reliability levels (El Geneidy et al, 2010) and therefore the modelling that follows will test whether this is the case in the context of the London Underground. There is likely to be strong correlation between this measure and the distance variable

**Intermediate Demand** – This variable is based upon identifying the intermediate stops on a given trip. For each of these stops an average demand is retrieved from demand calculated previously using the gate counts. The average demand at each of

these stations is then summed to give an average intermediate demand for the whole trip. It is hypothesised that greater demand between the origin and destination stations will negatively affect dwell times and contribute to decreasing reliability.

**Transfer** – This is a dummy variable similar to that specified by Uniman (2009). It takes a value of “1” if a transfer is necessary to complete a trip. If a trip can be made without a transfer, it is assumed that all travellers would not take another route where a transfer would be required and therefore a “0” is recorded.

**Deep line** – This is a dummy indicator that differentiates between two types of London Underground line: shallow and deep. Deep lines (e.g. Victoria) require more vertical travel between the surface and platform levels on the part of the passenger which may impact on reliability. These trips are marked with a “1”. Trips using a shallow line (e.g. Metropolitan) only are marked with a “0”. Where transfer trips use both a shallow and deep line, these are considered to be a deep line trip.

**Lift No Escalator** – This is a dummy variable used to denote whether a trip origin or destination is at a station where there are no escalators to transfer passengers between the surface and platform levels and a lift is provided instead (denoted “1”). It is hypothesised that using a lift will result in more variable travel times.

**Actual Journey Time – Scheduled In-Vehicle Time** – This is the actual average journey time observed in the Oyster data minus the scheduled in-vehicle time specified by Transport for London. This value is expected to be positive as all passengers have to access and egress to and from the platforms in addition to in-vehicle time. It is therefore a measure of these extra activities as well as average in-vehicle delay for a given OD. This measure will be referred to as the ‘congestion’ variable for the remainder of the chapter.

**Average Stop Spacing** – This is the first of the scaled variables, calculated by dividing the total distance between origin and destination with the count of intermediate stops (both are defined above). This overcomes the issue of number of stops being highly correlated with both distance and travel time.

**Average Demand at Intermediate Stop** – This is the second of the explanatory variables. As the number of stops is likely to be correlated with distance and travel time, so too is the total demand for an OD. To overcome this issue, the total intermediate demand between an origin and destination (i.e. the summation of demand at each intermediate stop) is divided by the total number of intermediate stops in order to calculate this variable.

**Congestion (scaled)** – The final scaled variable is the above congestion variable, divided by the average travel time. This is proposed as a candidate explanatory variable due to the scale of congestion likely being larger over longer duration trips. It therefore reduces correlation with the other distance and time based explanatory variables.

#### **4.4 Data**

This section will outline the process of preparing the 5% 2008 dataset in order to estimate a linear regression model with the aim of identifying predictors of reliability and forecasting reliability levels.

The starting point for the analysis was a SPSS dataset containing all 12,883,448 records of the 5% 2008 dataset (Dataset 1 from Chapter 3). The analysis was limited to the weekday AM peak and the London Underground mode only, although this methodology can be reproduced straightforwardly for other time periods, geographies and (to some extent) modes. Trips with very short travel times, the same origin and destination point, those that were incomplete and duplicate records were removed, resulting in 667,406 remaining records of individual trips. This dataset was imported into Microsoft Excel for the remainder of the data preparation step.

For each of the records, the origin and destination stations were identified. A count of the number of trips by origin and destination was then made, thereby allowing OD pairs with a sufficient sample size to be identified. For OD pairs without a transfer, this threshold was set at 100 records and resulted in 253 OD pairs, representing all London underground Lines (except Waterloo and City). Twenty further OD pairs including transfers were manually added to the dataset afterwards, although the sample size threshold for these ODs was 30 records.

Summary statistics for continuous variables that will be used for model estimation is produced in Table 4.2.

**Table 4.2 - Summary statistics calculated across ODs**

	<b>Average</b>	<b>Min</b>	<b>Max</b>	<b>St Dev</b>
<b>Travel Time</b>	23.29	4.39	48.40	8.89
<b>Standard Deviation Travel Time</b>	3.74	1.13	12.28	1.78
<b>Observations</b>	173.95	32	860	111.56
<b>Distance (km)</b>	9.00	0.86	34.95	5.89
<b>Intermediate Stops</b>	7.13	0	27	4.71
<b>Headway</b>	3.14	1.85	13.50	1.83
<b>Origin Demand (1 hour)</b>	227.85	22.15	833.66	224.13
<b>Destination Demand (1 Hour)</b>	152.81	3.85	1012.89	199.95

Table 4.2 shows that the method used to select OD pairs has provided a broad range of OD pair characteristics. This is demonstrated by the range of mean average travel times, which is 44 minutes. The minimum number of observations for an OD is 32, which is low in light of the size of the original dataset, but necessary to include enough transfer OD pairs. The demand at origin and destinations can vary substantively, although the averages presented in Table 4.2 are likely to be higher than the average for the whole London Underground network due to oversampling of busy OD pairs.

An issue briefly identified above is the possibility of strong correlation between distance, mean journey time and number of stops. This correlation among explanatory variables can lead to multicollinearity in regression models, potentially affecting the accuracy of parameter estimates and standard errors. Perfect collinearity is a violation of one of the assumptions upon which the Classical OLS regression model is based. To identify occurrences of high correlation between explanatory variables, the regression dataset was imported into SPSS and a correlation matrix of all variables was produced using the Pearson measure of correlation. This process also informed identification of the explanatory variables that were likely to be significant predictors of reliability levels. The correlation matrix is provided in Table 4.3

**Table 4.3 - Correlation matrix of initial variables**

		Mean_JT	STDev_JT	Headway	Origin_Demand	Destination_Demand	Transfer	Distance	Deepline	Lift_No_Escalator	Intermediate_Stops	Intermediate_Demand	MeanJT - SchedJT
Mean_JT	Correlation		0.489	0.479	-0.356	-0.174	0.257	0.776	.089	0.185	0.795	0.68	0.67
	Sig.		.000	.000	.000	.004	.000	.000	.143	.002	.000	.000	.000
STDev_JT	Correlation	0.489		0.376	-0.116	.051	0.278	0.435	-0.004	.115	0.345	0.359	0.525
	Sig.	.000		.000	.057	.401	.000	.000	.953	.057	.000	.000	.000
Headway	Correlation	0.479	0.376		-0.209	-0.021	0.491	0.498	-0.026	0.283	0.531	0.225	0.49
	Sig.	.000	.000		.001	.729	.000	.000	.668	.000	.000	.000	.000
Origin_Demand	Correlation	-0.356	-0.116	-0.209		0.564	-0.101	-0.384	-0.028	-0.209	-0.367	-0.215	-0.052
	Sig.	.000	.057	.001		.000	.100	.000	.650	.001	.000	.000	.423
Destination_Demand	Correlation	-0.174	.051	-0.021	0.564		0.328	-0.255	-0.054	.101	-0.247	-0.095	.125
	Sig.	.004	.401	.729	.000		.000	.000	.380	.098	.000	.120	.051
Transfer	Correlation	0.257	0.278	0.491	-0.101	0.328		.104	.093	0.633	.085	.096	0.433
	Sig.	.000	.000	.000	.100	.000		.086	.125	.000	.162	.113	.000
Distance	Correlation	0.776	0.435	0.498	-0.384	-0.255	.104		.069	.041	0.864	0.553	0.51
	Sig.	.000	.000	.000	.000	.000	.086		.254	.500	.000	.000	.000
Deepline	Correlation	.089	-0.004	-0.026	-0.028	-0.054	.093	.069		.048	.035	0.192	0.195
	Sig.	.143	.953	.668	.650	.380	.125	.254		.430	.560	.001	.002
Lift_No_Escalator	Correlation	0.185	.115	0.283	-0.209	.101	0.633	.041	.048		.046	.049	0.244
	Sig.	.002	.057	.000	.001	.098	.000	.500	.430		.445	.422	.000
Intermediate_Stops	Correlation	0.795	0.345	0.531	-0.367	-0.247	.085	0.864	.035	.046		0.696	0.402
	Sig.	.000	.000	.000	.000	.000	.162	.000	.560	.445		.000	.000
Intermediate_Demand	Correlation	0.68	0.359	0.225	-0.215	-0.095	.096	0.553	0.192	.049	0.696		0.394
	Sig.	.000	.000	.000	.000	.120	.113	.000	.001	.422	.000		.000
MeanJT - SchedJT	Correlation	0.67	0.525	0.49	-0.052	.125	0.433	0.51	0.195	0.244	0.402	0.394	
	Sig.	.000	.000	.000	.423	.051	.000	.000	.002	.000	.000	.000	

In Table 4.3 any correlations outside the range -0.5 to 0.5 are highlighted in red. These high positive or negative correlations between explanatory variables are likely to be sources of collinearity. Also highlighted in bold is the row for the dependent variable ‘STDev\_JT’ (the standard deviation of journey time for an OD). Variables that correlate highly with the standard deviation of journey time are likely to be good predictors in the course of model estimation.

Table 4.3 shows that mean journey time, distance, intermediate stops and intermediate demand are highly correlated with each other. The key relationship is between time and distance, which would be expected to have a strong positive correlation. Distance between origin and destination is also positively correlated with the number of stops. As intermediate demand is additive along a route, a positive correlation between intermediate demand and number of stops would also seem intuitive.

Of further interest is the correlation between mean journey time and mean journey time minus scheduled journey time, where the latter is a measure of congestion. This is because the scale of congestion has the potential to be larger over longer journey times. This result suggests there will be significant collinearity between the number of stops, intermediate demand and the congestion measure. One possible solution to this issue is to make use of the three scaled variables outlined at the end of Section 4.3.

Standard deviation of journey time will be the indicator of reliability used in the regression models that follow. Table 4.3 shows strong and statistically significant correlation between this variable and mean journey time, headway, transfer, distance, intermediate stops, intermediate demand and congestion. These variables are therefore likely to be predictors of reliability levels, although all variables will be included in initial model estimation in the next section.

## **4.5 Model Estimation**

The models estimated in this section are run using the OLS linear regression procedure in SPSS. The dependent variable in all cases is the standard deviation of journey time. The forward stepwise procedure is used, where each iteration adds a

significant variable from those specified until an optimal model is found. The adjusted R<sup>2</sup> will be reported and is interpreted as indicative of the performance of the model.

**Model 1: Full Model**

This model initially included all variables presented in Table 4.3. It therefore takes the following form:

$$\begin{aligned}
 Reliability_{i,j} = & \beta_0 + \beta_1 Mean\_JT_{i,j} + \beta_2 Headway_{i,j} + \beta_3 Demand_i & (4.3) \\
 & + \beta_4 Demand_j + \beta_5 Transfer_{i,j} + \beta_6 Distance_{i,j} \\
 & + \beta_7 Deepline_{i,j} + \beta_8 Lift\_no\_escalator_{i,j} \\
 & + \beta_9 Intermediate Stops_{i,j} \\
 & + \beta_{10} Intermediate Demand_{i,j} \\
 & + \beta_{11} (Mean JT - Sched JT)_{i,j}
 \end{aligned}$$

The stepwise procedure excludes the majority of the explanatory variables: mean journey time, origin demand, destination demand, distance, deepline, lift no escalator and intermediate stops. This leaves the congestion variable, transfer and headway as forming explanatory variables, with a model adjusted R<sup>2</sup> of 0.366. The parameter estimates are presented in Table 4.4.

**Table 4.4** - Parameter estimates of Model 1

	Beta	t
(Constant)	.904	2.873
MeanJT_SchedJT	.182	3.834
Transfer	1.239	2.778
Headway	.242	3.178

The signs of the parameter estimates in Model 1 are in line with expectations: higher levels of congestion, the existence of a transfer, and higher headways would all result in an increased standard deviation of travel time. The model is less convincing insofar as there is no distance or time variable represented. This may be a result of the correlation between congestion and journey time/distance. This will be addressed in Model 2.



### Model 2: Full Model (Scaled Variables)

The same model is run again, but replacing the number of stops with average stop spacing, intermediate demand with average demand at intermediate stop, and congestion is scaled with average travel time. This should reduce the amount of collinearity within the model and therefore produce a statistically significant estimate of mean journey time in line with much of the previous work.

**Table 4.5** - Parameter estimates of Model 2

	Beta	t
(Constant)	1.687	2.764
Mean_JT	.073	6.019
Demand_Stop	.001	2.721
Congestion_Scaled	2.349	2.413
Stop_spacing	-.646	-2.446
Deepline	-.580	-2.213

Model 2 estimates parameters for two of the three scaled variables described above. Congestion remains a positive predictor of reliability levels. Stop spacing, measured in km, has a negative relationship with reliability, which is an intuitive result implying that underground lines with greater stop spacing will be more reliable. Mean journey time is now a significant variable, which is a finding of the aforementioned study of Uniman (2009). The demand / stop parameter is positive, implying that greater average demand at intermediate stops between an origin and destination will decrease reliability. The deepline parameter estimate would be expected to be positive as it would require greater movement within the origin and destination station; a negative result implies that trips on deepline underground lines are more reliable than those close to the surface. The adjusted  $R^2$  of Model 2 is 0.183, which implies a poorer overall predictive capacity than Model 1. This low  $R^2$  is caused by a wide variance of observations from the mean, as illustrated by the scatter plots in Appendix 1. A further exploratory model run shows that this  $R^2$  would be further reduced when mean journey time was replaced by distance.

## 4.6 Discussion

The models presented in Section 4.5 demonstrate a broad range of factors impacting on reliability of the London Underground. In both Model 1 and Model 2 it is congestion which is a key explanatory variable (both scaled and unscaled by journey time). Other factors found to be significant are transfer, headway, mean journey time, demand per intermediate stop and stop spacing. A key issue identified was the correlation between distance based variables. The impact of this can be observed in Model 1, where the congestion variable accounted for the entire distance based impact. When this was accounted for in Model 2 (by scaling the congestion variable), mean journey time became a key driver of reliability levels. Table 4.3 showed that mean journey time and distance were highly correlated ( $\rho > 0.75$ ) which accounts for why parameters for these variables were not estimated alongside one another in a single model.

The key finding of the chapter is that a broad range of factors impact upon reliability levels of the London Underground, and there appears to be some broad agreement of the causal factors which include demand and travel time. The categorical variable *lift no escalator* was not found to be a substantive driver of reliability. The *Deepline* categorical variable was found to be statistically significant and negative. There is no obvious interpretation to this and it does not strongly correlate with any other explanatory variable.

Given the relatively low explanatory power of the two models presented in this chapter, the conclusion is made that whilst they are useful for identifying the factors affecting reliability, they could not be used with confidence to predict future levels of reliability.

## **4.7 Conclusion**

The model estimation above contributes to Objective 2 of this thesis by demonstrating that the Oyster card (and similar systems) can provide data capable of revealing the factors affecting reliability on public transport. It also meets Objective 3 by developing the means to better understand the factors affecting reliability.

The work builds upon that conducted previously by other researchers, which was primarily in the context on bus operation. It also builds upon the analysis of Uniman (2009) (who also used Oyster card data) by utilising a fuller range of explanatory

variables. The explanatory power of the model proposed by Uniman was greater than any estimated in this chapter, but this can likely be attributed to the use of an ‘incident’ variable which was calculated based upon reliability levels observed. The next step would be to estimate similar models based upon other public transport modes, including bus.

# Chapter 5 - Discrete Choice Modelling with Smartcard

## Datasets

### 5.0 Introduction

The previous chapter was an investigation into the factors driving the reliability of public transport. This chapter will combine Mean-Variance (MV) with the smartcard datasets to obtain parameter estimates related to passenger demand for public transport. These parameter estimates will allow a Reliability Ratio (RR) to be calculated. The method developed and applied in this chapter will be related to London's public transport network. It will however be sufficiently broad to be applicable to similar public transport networks where smartcard data are available.

It was clear from Chapter 2 that travel time variation (TTV) is an important factor for explaining agents' travel behaviour (Eddington, 2006), but that its valuation remains an active research strand. There has been a general reliance on Stated Preference (SP) studies in order to value TTV (Ettema and Timmermans, 2006; Batley and Ibáñez, 2012; Börjesson et al, 2012). Meta-analysis of these SP studies (e.g. Li et al, 2010; Carrion and Levinson, 2012), as well as expert workshops (e.g. de Jong et al, 2009), have been only partially successful in clarifying the value of TTV. It has been noted that a Revealed Preference (RP) methodology may be instructive in the valuation of TTV, but the situations where it can be effectively applied are rare (Bates et al, 2001), and consequently this area remains underdeveloped. This chapter is an attempt to treat the Oyster Card data as an RP dataset. This will enable estimation of a Reliability Ratio (RR) when combined with a standard discrete choice model.

The specific contributions of the present chapter are:

1. To develop an RP methodology for the valuation of TTV using Oyster data
2. To utilise automatically generated datasets to provide valuation evidence for TTV
3. To demonstrate the use of such a dataset with existing discrete choice modelling techniques
4. To evaluate the effectiveness of the methodology developed.

The chapter is arranged as follows. In Section 5.1 SP and RP are re-introduced from Chapter 2. The differences between RP and SP-based methods are outlined and the potential for RP to offer an improvement over SP is discussed. In Section 5.2, the approach to modelling reliability is outlined and the relevant discrete choice model specifications are introduced. In Section 5.3 the method of selecting and applying the 2011 5% sample dataset (designated Dataset 2 in Chapter 3) is outlined. Issues with the dataset are addressed, including omitted modes and limiting assumptions. The data are then applied with choice modelling methods and basic MV model specification in Section 5.4. The models are estimated based upon choice between alternatives containing one or more rail based modes. These models and the resulting RRs are discussed and the success of the method is reflected upon in the discussion. In the concluding section, the use of public transport smartcard data as an RP data source is reflected upon in light of the method developed and the results obtained through discrete choice modelling of the data.

## 5.1 Background

### 5.1.1 Travel Time Variation (TTV)

The approach for modelling reliability in this chapter is MV. This approach allows for the estimation of a RR, which can be benchmarked against values found in the literature. It does however come with an assumption that every available travel option can be accurately estimated and utilised in travellers' decision making processes. To recap from Chapter 2, the RR is given by the ratio of the marginal utility of the standard deviation of travel time to the marginal utility of mean travel time:

$$RR = \frac{\frac{\delta U}{\delta \sigma}}{\frac{\delta U}{\delta \mu}} \quad (5.1)$$

Where  $\frac{\delta U}{\delta \mu}$  is the marginal utility of the mean travel time  $\mu$  with respect to utility  $U$ , expected to be negative as travel time is usually treated by travellers as bad.  $\frac{\partial U}{\partial \sigma}$  is the marginal utility of the standard deviation of travel time  $\sigma$  with respect to  $U$ , also

expected to be negative as travellers are generally risk averse as shown in Chapter 2. The value of RR is therefore expected to be a positive value.

The literature review of RR studies in Chapter 2 showed that a broad range of RR estimates have been made. Further evidence of this phenomenon is available in the meta-analyses of Li et al (2010), Carrion and Levinson (2012) and Wardman and Batley (2014). The meta-analysis conducted in Chapter 2 could not find strong evidence that the mode, method of data collection or reliability framework employed would systematically affect the RR estimate obtained.

In addition to this issue, the literature review outlined the well documented drawbacks of hypothetical choice questionnaires, summarised in a recent review of travel time reliability (Wardman and Batley, 2014). This paper suggested that a strategic bias may be observed given the contentiousness of TTV or lateness to travellers, particularly when the purpose of the study is clear to the respondent. Furthermore, it might be added that the general difficulties related to SP of misapprehension, fatigue and boredom experienced by respondents might be exacerbated when dealing with the complexity of TTV issues.

The wide range of RR estimates found in the aforementioned meta-analyses suggest that it would be appropriate to calculate a unique RR for each context for which it was required. This approach has practical implications if SP-based studies were used to estimate a RR; these being the cost, resource and analytical knowledge required in every case. A possible alternative approach for practitioners is to ‘transfer’ a pre-existing RR which was calculated under similar circumstances using SP (i.e. mode, geographic location etc.) to the new context. It is clear however that an appropriate RR will not be available for all situations. This does not take into account issues around context and fungibility (Orr et al, 2012). Nevertheless, the utilisation of a RR from another context is often the current position of practice: see WebTAG Unit A1.3 for evidence of this from the UK. The reason for this situation is a tacit acknowledgement of the difficulty and resources required for conducting a unique choice experiment for each occasion that a RR is required.

It is clear then that accurate estimates of RRs are necessary, but that current methods to estimate it raise questions over their suitability. An alternative or complementary solution to the estimation of RRs might be appropriate; one which overcomes:

- how to make the appropriate choice of RR from the broad range of values available in the literature
- issues of SP respondent understanding of TTV
- presentational issues specifically related to TTV SP surveys
- general problems related to SP (strategic/protest responses)
- cost/resource required to estimate a local RR

To address these issues, attention is now turned to the less well researched area of RP based estimates of the RR.

### **5.1.2 The Revealed Preference approach**

The review of RP studies in Chapter 2 was based upon what actual traveller choices can tell the analyst about traveller preferences. In the context of transport reliability, this is how they might in practice trade-off travel time, TTV and other attributes of the journey. An RP based methodology would overcome the issues highlighted in the previous section by incorporating participants' actual decision making choices, including imperfections such as habitual behaviour and imperfect information availability (Wardman and Batley, 2014). An RP only approach could utilise readily available datasets and has the potential to be largely automated. This approach could therefore provide a cost-effective, rapid, temporal/location specific method of estimating a RR.

The general concerns with an RP-only approach are that insufficient situations might exist in a dataset for effective estimation of parameters, as well as difficulty for the analyst in knowing the completeness of travellers' information. This latter concern includes imposing choices upon individuals by the analyst, and difficulty in specifying a realistic choice set. These issues are not relevant in SP contexts where knowledge can be considered perfect. The literature also notes a difficulty specific to MV applications where travel time and standard deviation of travel time have been found to be correlated (Batley et al, 2008; de Jong et al, 2009). An additional criticism of RP based methods

might be the restriction to situations where appropriate data are available (Bates et al, 2001). This chapter will demonstrate that this concern will become less of a problem as automatic data collection systems become more prevalent.

## **5.2 Modelling Framework**

### **5.2.1 Choice and Risk**

The phrase ‘choice under uncertainty’ is often used in the literature on travel time variation, however in this chapter ‘choice under risk’ is used. This distinction is vital: ‘choice under risk’ denotes a situation where the distribution of outcomes is known by the traveller – a key assumption for the methodology. Although unrealistic, this is a simplifying assumption which will allow a RR to be estimated based upon RP data. Further work (in particular a longer temporal duration of data) could take into account learning or knowledge effects. The axioms of Expected Utility (EU) theory are applied as an account of the observed traveller behaviour and an introduction to this can be found in Chapter 2. The literature has shown that the EUT approach can provide results at odds with actual human behaviour (Kahneman and Tversky, 1979), but nevertheless it is accepted in the transportation literature as a useful tool for understanding risk. In accepting von Neumann and Morgenstern’s EU framework and implementing this through a discrete choice model, each of the travellers observed is assumed to have well defined, complete and transitive preferences.

### **5.2.2 Expected Utility**

The EU function for the estimation of a RR using the MV framework is outlined in Chapter 2 of this thesis.

In what follows, the MV framework for travel time variation will form the basis of modelling passengers’ behavioural choices. The MV-based expected utility function takes the following general form:

$$EU_{MV} = \beta_1\mu + \beta_2\sigma \quad (5.2)$$



Where  $\mu$  is the mean travel time and  $\sigma$  is the standard deviation of travel time, and  $\beta_1$  and  $\beta_2$  are preference parameters for these variables.

### **5.2.3 Data Requirements**

The RP estimation of a RR will be based upon travellers' choice responses to observable supply side metrics of a public transport system. A requirement of this method is a dataset which can provide estimates of mean and standard deviation of travel time and records of travellers' mode choices. The smartcard datasets described in Chapter 3 meet these criteria. It may be possible to apply this method in contexts outside public transport; in the context of private car travel routing and travel time, datasets may be obtainable from the providers of satellite navigation systems or operators of automatic number plate recognition (ANPR) systems.

There are a number of dimensions that could be used to form the choice situation. Examples include departure time, route choice, and mode choice. In the work that follows, the choice for individuals will be between alternatives that are based upon modes. In the case of departure time and route choice, insufficient information is available in the dataset to identify alternatives. However, Chapter 3 identified that the mode (or combination of modes) used to complete a journey would be identified via the SUBSYSTEMID field. OD pairs where a number of SUBSYSTEMID identifiers have been used across all travellers will form the basis of the choice.

For this method, it is important initially to identify situations where all possible mode based alternatives between an origin-destination pair can be observed, and suitable sample sizes are available for each choice. There must also be some means of knowing which mode is used – through intermediate observation or some indication at the origin or destination point. Section 5.3.1 will describe the process of developing the choice dataset from Dataset 2 described in Chapter 3.

### **5.2.4 Choice Model Specification**

This methodology will make use of common discrete choice model specifications to represent the mode choice of the traveller. These will include the multinomial logit (MNL) and cross-nested logit (CNL) models. Both model specifications assume that the

decision maker has a number of alternatives from which to make their choice, called the *choice set*. Train (2009) sets out the three key properties of the choice set; that it is:

- Exhaustive – it contains all possible alternatives that could be chosen
- Mutually exclusive – that only one alternative can be chosen
- Finite

In line with EU theory, the decision maker will assign an expected utility to each of the available alternatives, and then choose the alternative with the highest expected utility. The expected utility for each alternative is determined by attributes of the alternative as well as attributes of the decision maker (Ortúzar and Willumsen, 2011). That these attributes usually cannot be known accurately has led to the development of Random Utility Models (RUMs), of which MNL and CNL are members. These models contain a random error term for each individual and choice, which is reflection of:

- The subjectivity of decision makers (including limits to rationality)
- Omission in model specification of relevant variables
- Misspecification of the model
- Error in data collection

The EU function within the context of RUMs therefore consists of two main elements: that which is observed for a given individual  $n$  for alternative  $j$ , usually denoted by  $V_{nj}$ , and a corresponding unobserved element, often referred to as the *error term*, is denoted by  $\varepsilon_{nj}$ . This latter element is usually unknown to the researcher, and is therefore randomly drawn from a distribution. It is the shape of this distribution, defined by the researcher, which defines the type of model that is being estimated. For example, a Normally distributed error term results in the probit model. McFadden (1974) specified the unobserved utility as type 1 extreme value, and derived the commonly used logit model.

The multinomial logit (MNL) model is among the simplest and most well-known choice model, where the probability of individual  $n$  choosing a given alternative  $i$  ( $P_{ni}$ ) from choice set  $J$  ( $i \in J$ , where  $j = 1, \dots, J$ ) is given by:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \quad (5.3)$$

Where the observed portion of utility for individual  $n$  and alternative  $j$  ( $V_{nj}$ ) is linear in parameters and is therefore given by the term  $\beta' x_{nj}$  (where  $x_{nj}$  is a vector of observed variables for individual  $n$  and alternative  $j$ ).

A key assumption is the condition of independence from irrelevant alternatives (IIA) between the error terms of alternatives which is often not satisfied in real-world applications; there is usually correlation between the error terms. This issue can be overcome by either a mixture or nested logit model specification.

The nested logit (NL) model can be specified in most choice modelling software packages; similar alternatives are explicitly defined to be contained within a single nest. A common example of this is a public transport nest containing bus and rail and a separate nest containing only private car. This model specification could therefore acknowledge that the public transport modes are closer substitutes for one another than a single public transport mode and the private car.

An issue with this simple nested structure is that modes may have membership of more than one nest. In the case of London's transport system, travellers may choose to make their trip through a combination of modes. For example, a choice set for travel between two points may be between using a combination of the London Underground and the Docklands Light Railway (DLR), or the Underground and Heavy Rail services. In this case both modes belong to the Underground nest, but the first such mode would also belong to a DLR nest, and the second to a Heavy Rail nest. The natural choice model specification for this situation is the cross-nested logit (CNL) model.

It was Bierlaire (2006) who provided formal proof that the CNL model was consistent with RUM. Wen and Koppelman (2001) presented the generalised nested logit (GNL) model and showed how other common discrete choice models were special cases of it. The CNL model takes the following form:

$$P_{ni} = \frac{\sum_{k=1}^K (\alpha_{ik} e^{V_{ni}})^{\frac{1}{\lambda_k}} \left[ \sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{\frac{1}{\lambda_k}} \right]^{\lambda_k - 1}}{\sum_{k=1}^K \left[ \sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{\frac{1}{\lambda_k}} \right]^{\lambda_k}}. \quad (5.4)$$

The subscript  $k$  represents an indexation of the nests in the model, with  $B_k$  denoting a nest of alternatives.  $\alpha_{ik}$  is the allocation parameter which signifies the proportion of membership of alternative  $i$  to nest  $k$ . This model is general, so that each alternative can belong by varying degrees to any number of available nests. The restriction is imposed that:

$$\sum \alpha_{ik} = 1 \quad (5.5)$$

If the allocation parameter for an alternative  $i$  in nest  $k$  is equal to one, then the alternative belongs wholly to nest  $k$ . Equally, an allocation parameter of zero suggests that the alternative  $i$  is not associated with nest  $k$ . If all allocation parameters are either one or zero, this would imply that the model structure is that of *nested logit*. This is why the model presented in Equation 5.4 has also been named the *Generalised Nested Logit* (Wen and Koppelman, 2001).

Both MNL and CNL models will be utilised in the modelling section of this chapter.

## 5.3 Empirical Application of the Model

### 5.3.1 Smartcard Data

Chapter 3 introduced smartcards as a subject of research and demonstrated that public transport smartcards provide a unique research opportunity in their own right. Researchers have recognised the value of smartcards in observing travel behaviour (Choi et al, 2012) and monitoring system performance (Jang, 2010), but there is less research available which models both supply and demand sides of a transport system.

The data available for estimating the choice model were the trips made by a 5% sample of Oyster Card users on the Transport for London (TfL) public transport network during a single, non-school holiday month in 2011 (Dataset 2 in Chapter 3). The 2011 5% was preferred to the 2008 data due to the larger number of records in the

former. For the purposes of estimating the discrete choice models, the analysis was restricted to the AM peak period. Analysis in Chapter 3 showed that a large proportion of demand would occur during this period, which therefore increases the number of origin-destination (OD) pairs with adequate sample sizes. Restricting analysis to the AM peak also results in relatively homogenous scheduled travel times for a given OD pair, thereby reducing the impact of varying schedules throughout the day on travellers' journey times. Finally, analysis in Chapter 3 suggested that the AM peak would consist of the largest number of commuters; and it is likely this group of travellers would have a higher level of knowledge of reliability levels between their origin and destination than leisure travellers. For these reasons, non-AM peak journeys (beginning outside 07:00 – 10:00) and weekend journeys were removed. Bus trips were also removed due to difficulty in accurately identifying the time that the passenger alighted from the bus. Including this mode would require an additional dataset detailing individual bus performance, which was not available to this study.

To estimate the parameters of the discrete choice models it was necessary to identify the OD pairs where a choice between modes was observable. This was done by creating an OD trip matrix specific to each of the modes available. A database relationship was created between each of these matrices to identify ODs where two or more modes were used with a count of at least 15 observations per mode. This process resulted in 74 candidate OD pairs for the AM peak model estimation, giving a total of 9,408 usable records (an average of 145 trips per OD). However, an issue was identified of dominated OD pairs; defined as ODs where both the mean travel time and standard deviation of travel time were lower on one of the two mode choices available. On such OD pairs there would be no clear trade-off between travel time and travel time risk which would affect estimation of the model parameters. Consequently, these OD pairs were removed from the dataset. This process reduced the number of OD pairs from 74 to 37, and the number of records in the choice dataset from 9,408 to 4,918. As a result of these processes, two modes (Light Rail and Light Rail/Heavy Rail) had only three choice situations (i.e. ODs) each. This meant that these modes could not be accurately modelled and therefore they were removed from the dataset. The final dataset used for model estimation therefore had 31 ODs, made up from 4,140 Oyster records. The

number of records available at each step is shown in Table 5.1, alongside the proportion of the total dataset.

**Table 5.1** - Number of records available at each stage of analysis

	<b>Count of records</b>	<b>% of total</b>
Full 5% Dataset	22,804,496	100.00%
Remove Bus	7,676,570	33.66%
Remove Unfinished/started trips	6,851,665	30.05%
Remove Short Trips	6,834,036	29.97%
Remove Duplicates	4,605,511	20.20%
Remove Weekend	3,454,133	15.15%
Remove non AM Peak	1,001,134	4.39%
Two modes for OD	9,408	0.04%
Remove Dominated OD	4,918	0.02%
Insufficient Alternatives	4,104	0.02%

Table 5.1 shows that the final choice dataset contained only 0.02% of the original 22,804,496 records in the 2011 5% dataset. Losses of data occur at every step, but key elements include the removal of the bus mode (a loss of approximately two thirds of the records), the removal of duplicates, and finally the requirement of two modes with at least 15 records for a given OD. This demonstrates the requirement to begin analysis with large and/or targeted data in order to ensure that the choice dataset is sufficiently large.

It was also necessary to identify if bus travel was occurring on any of these 31 ODs, creating bias within the choice models by making the PT choice set non-exhaustive. To address this concern, a list of all origin and destination stations in the choice dataset was created. Returning to the full dataset of 11 million records, it was possible to identify all travellers that interacted with these origins and destinations, including trips made using bus for all or part of the trip. Drawing on the work of Seaborn et al (2009) it was then possible to link transfers between bus and other modes, and thereby identify instances of travellers using bus for all or part of their trip between any of the OD pairs in the choice dataset. Some evidence was found of a small number of travellers using bus between five of the ODs, however the numbers were small (one or two travellers for each of these OD pairs). On reflection, it was decided to proceed to the analysis stage

with all 31 AM peak OD pairs. An alternative procedure would have been to exclude the OD pairs where there was evidence of bus use, but much data would have been lost.

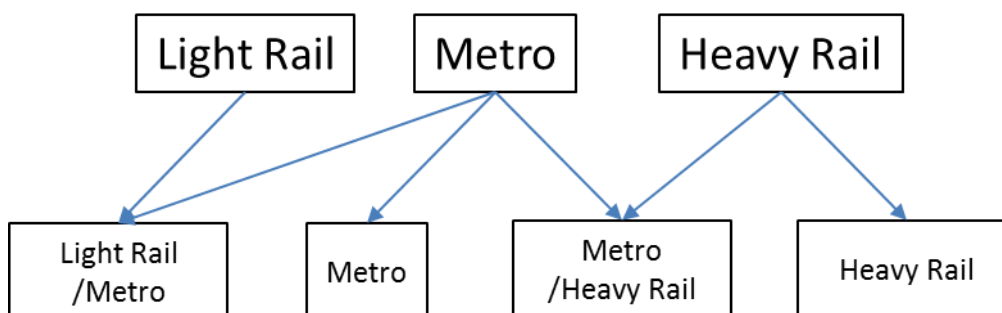
A related issue is that no highway based modes were modelled – including private car. The assumption was made that the car mode would notionally exist in an independent ‘non-public transport’ nest, and that the estimates within the public transport (PT) nest were therefore unbiased.

### 5.3.2 Data Analysis

. Within this dataset, the use of three public transport modes was identified:

- Light Rail – also referred to as the Docklands Light Railway, or DLR
- Metro – also referred to as London Underground, or LU
- Heavy Rail – standard heavy rail services, including London Overground and referred to as ‘Rail’ in the model specification

Many journeys consist of two of these modes and therefore a maximum of six modes could have been modelled: Metro, Light Rail, Heavy Rail, Metro/Light Rail, Metro/Heavy Rail, Light Rail/Heavy Rail. However, recognising the problem of a low number of OD pairs for two modes, the analysis made use of only four of these modes. Figure 5.1 shows how each of the four alternatives was configured. The method of grouping recognised that there would be correlation between the four alternatives when they contained a common mode of transport. Therefore, the aforementioned CNL structure was an appropriate form for the choice model specification, which will be estimated in Section 5.4.



**Figure 5.1** - A mapping of rail based modes to alternatives used in subsequent modelling

For any one OD pair, there were only ever two of the four alternatives available. If a traveller entered a Metro gate at the origin station and existed at a Metro gate at the destination station, and the trip were possible using Metro only, then this would be designated as a Metro trip by TfL. If, however another trip began at a Metro gate at the origin and ended at a Heavy Rail gate at the destination, this trip would be designated Metro/Heavy Rail.

Following from the calculation of mean and standard deviation in Chapter 3, it was a simple process to now calculate a sample mean travel time and sample standard deviation by using both OD and alternative – i.e. each OD had two sets of MV values; one for each of the public transport travel choices available to the traveller, calculated across all travellers using that mode/OD combination.

Each individual record, representing a single trip by a traveller, was assigned the mean and standard deviation for the two options available.

The previously estimated regression models showed that a range of explanatory variables would impact upon the levels of reliability. Nassir et al (2015) suggested that a range of variables, in addition to mean travel time and standard deviation of travel time, would impact upon traveller choices. Therefore, additional variables were calculated, parameters for which will be estimated in Model 1c below. These were:

- Transfer – a categorical variable indicating whether a transfer was required on a given alternative
- Headway – A calculation of average time between vehicles during the AM Peak. Where two lines or modes were used, the headways were added together.

A variable to represent station size was also intended to be used, but due to sample size issues all stations were large, and preliminary analysis showed that this variable would not be significant. Therefore the variable was not included in further analysis.

The summary statistics related to the four explanatory variables are presented here, with a further breakdown to the OD level in Appendix 2.

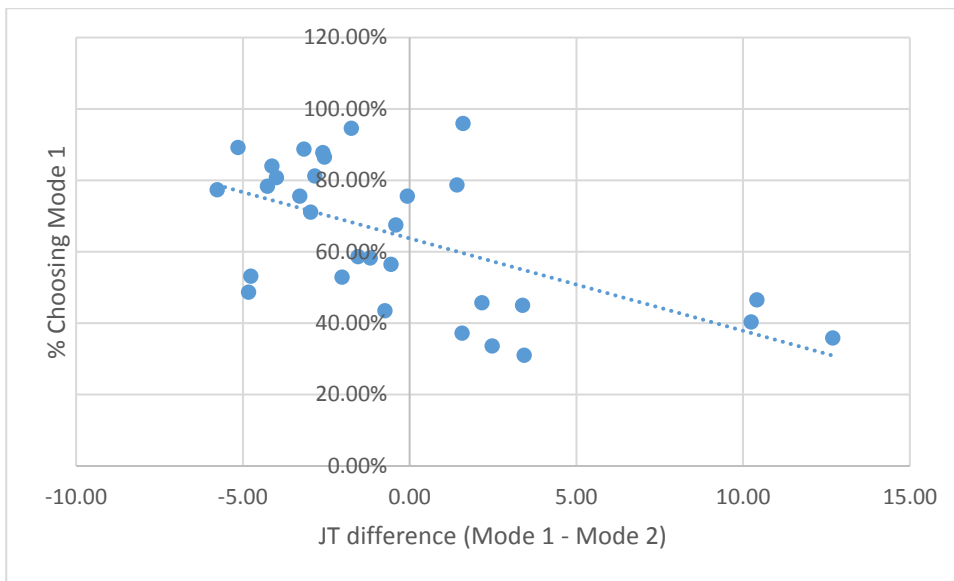


**Table 5.2** - Summary statistics of dataset used to estimate choice models

<b>Mode</b>	<b>Count</b>	<b>Average JT</b>	<b>Average St Dev</b>	<b>Transfer Trips</b>	<b>Average Headway</b>
LU	27	31.77	5.18	24	5.24
LU/DLR	18	40.67	4.36	18	7.73
Rail	6	28.84	5.36	5	17.46
LU/Rail	11	30.10	5.37	11	12.77
<b>Overall</b>	<b>62</b>	<b>32.68</b>	<b>5.10</b>	<b>58</b>	<b>6.86</b>

Table 5.2 summarises all 62 travel options available in the choice dataset (31 ODs, with two mode options each). The averages quoted are weighted by number of passengers on each mode. The LU mode is well represented, being available for 27 of the 31 ODs. Journeys on rail are, on average, the shortest duration, with LU/DLR the longest. This is unsurprising given that stop spacing and vehicles speeds tend to be higher on heavy rail in comparison to light rail. In contrast, the average standard deviation is lowest on the LU/DLR mode. The average headway is lowest on the LU mode; this is intuitive as many lines on the London Underground operate with only two minutes between vehicles during the AM peak.

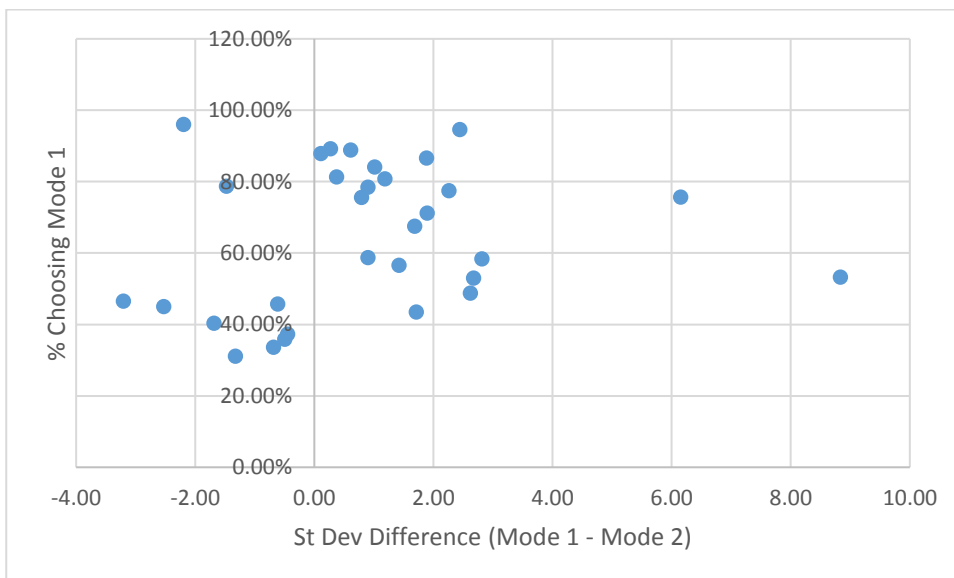
Using the values in Appendix 2 it is possible to plot the differences on a given OD against the number of passengers choosing an option. In the first example below, the difference between the average journey time for Mode 1 and 2 on an OD is plotted against the proportion of travellers using Mode 1. If passengers dislike travel time as expected, a relationship with a negative gradient would be observed: where travel times are higher on Mode 2, a greater proportion of travellers would be observed using Mode 1. This relationship can be observed in Figure 5.2.



**Figure 5.2 - Mean journey time difference plotted against mode choice**

Figure 5.2 shows, on average, fewer passengers using Mode 1 when its travel time is higher than that on Mode 2, and more passengers using Mode 1 when the travel time is lower. The line of best fits illustrates this trend, although a relatively low  $R^2$  of 0.35 would suggest that there may be other factors that influence mode choice.

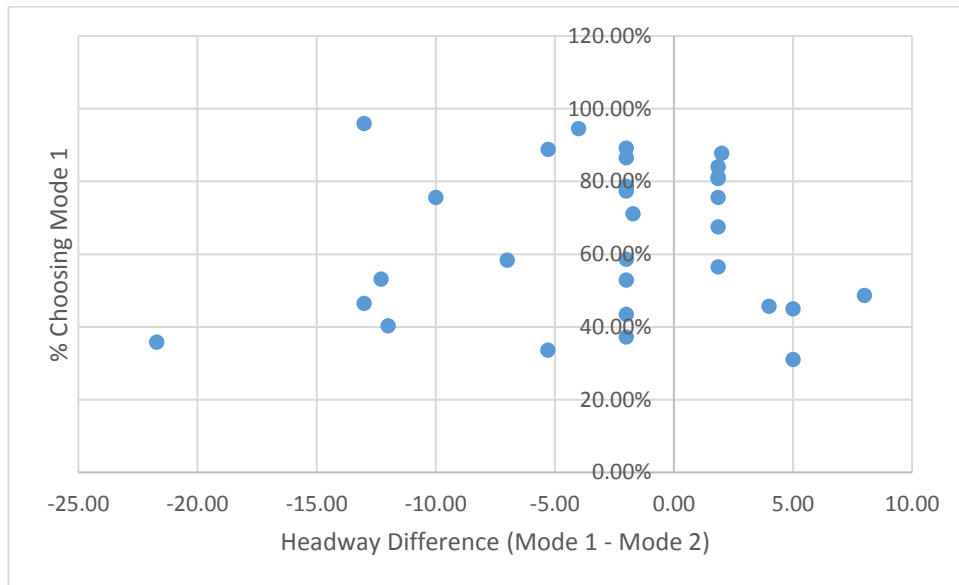
Figure 5.3 is a similar plot to 5.2, with the mean average travel time replaced by the standard deviation of travel time.



**Figure 5.3 - Standard deviation of journey time difference plotted against mode choice**

The relationship between relative standard deviation and mode usage is not clear from Figure 5.3. Seven of the nine negative points below zero involve the Heavy Rail

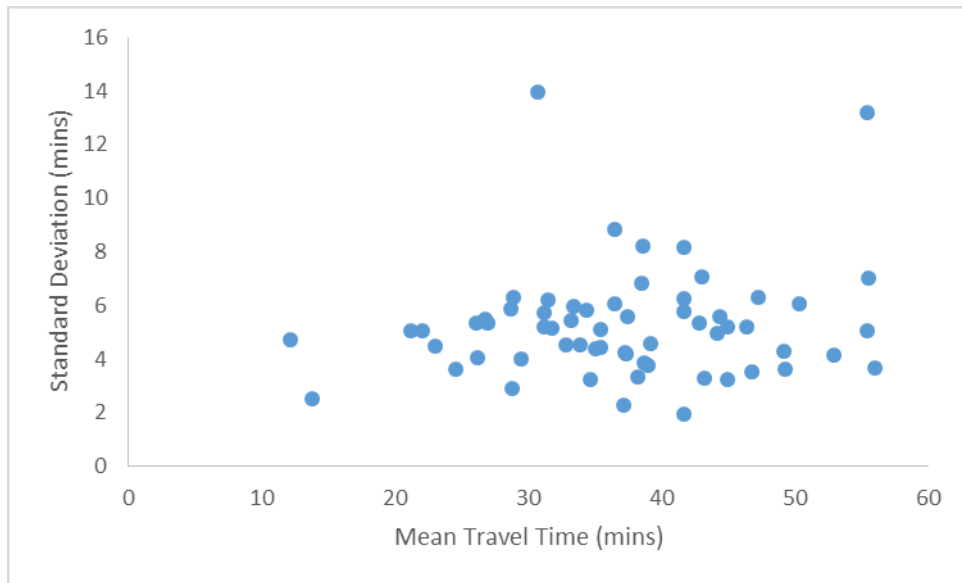
mode, meaning that some mode specific affects may be affecting this result. It will therefore be crucial to represent mode specific effects in subsequent model estimation.



**Figure 5.4** - Headway differences plotted against mode share

Figure 5.4 does not clearly demonstrate a relationship between headways and mode choice. A negative relationship would be expected under the hypothesis that passengers prefer a public transport mode that has a lower headway. Many OD pairs are observed with little difference between headways (i.e. close to zero on the x axis), and so it is unclear whether the choice models estimated subsequently will identify a statistically significant parameter associated with headway.

A concern remains regarding the issue raised in the literature of correlation between mean and standard deviation of travel time (Batley et al, 2008; de Jong et al, 2009), however the analysis based upon the information in Appendix 2 suggested that this was not an issue for these data. This relationship for all 62 OD/mode combinations is plotted in Figure 5.5.



**Figure 5.5** - Scatterplot of mean travel time against standard deviation of travel time for 62 OD/mode combinations

Figure 5.5 entails no statistically significant correlation between the two variables, and therefore the conclusion was made that this would not unduly affect the estimation of the choice models.

Chapter 3 noted that the journey time data included all activities that occurred behind the gate. Therefore there is likely to be correlation between journey time and the additional variables of headway and transfer. To investigate this possibility, a correlation matrix of all four variables was calculated based upon all 62 mode/OD combinations in the dataset.

**Table 5.3 - Correlation matrix of explanatory variables**

		MEAN_JT	ST_DEV_JT	TRANSFER	HEADWAY
MEAN_JT	Correlation	1	0.137	.282*	-0.222
	Sig. (2-tailed)		0.288	0.026	0.082
ST_DEV_JT	Correlation	0.137	1	0.031	0.027
	Sig. (2-tailed)	0.288		0.811	0.834
TRANSFER	Correlation	.282*	0.031	1	0.019
	Sig. (2-tailed)	0.026	0.811		0.883
HEADWAY	Correlation	-0.222	0.027	0.019	1
	Sig. (2-tailed)	0.082	0.834	0.883	

\*. Correlation is significant at the 0.05 level (2-tailed).

The correlation matrix in Table 5.3 shows positive and statistically significant correlation between mean journey time and transfers, which may affect model estimation. The relationship between headway and mean journey time would be expected to be positive as increased time between vehicles would increase travel time, all else equal. However the correlation above is negative, although this was not statistically significant. The standard deviation of journey time does not appear to correlate strongly with the transfer or headway variable, and therefore should not affect its parameter estimate substantively. Table 5.3 is further expanded in Appendix 3 with scatterplots showing the relationships between variables in greater detail (excluding the categorical transfer variable). The plots do not visually indicate any clear relationships between any combinations of variables.

Another issue prior to model estimation was the extent to which individuals were observed on multiple occasions. One reason that the dataset was restricted to AM peak was to capture commuters capable of estimating the travel time and standard deviation of travel time on different alternatives. This would therefore imply repeated journeys made by the same individuals. The table below shows the number of journeys made by the same Oyster card within the choice dataset. The assumption is made that each Oyster card was used by one individual only.

**Table 5.4 - Number of uses by Oyster card ID**

<b>Uses</b>	<b>Count of Oyster Cards</b>
1	410
2	71
3	44
4	35
5	22
6	21
7	19
8	16
9	10
10	10
11	20
12	17
13	10
14	20
15	23
16	15
17	16
18	23
19	16
20	11

Table 5.4 shows that there are 829 separate Oyster cards used within the choice dataset with an average of 5 transactions by each card. 410 of these cards register a trip only once over the study period, whereas just over 20% of cards are used more than 10 times. Although this does imply that many of the trips recorded in the dataset are made by those unfamiliar with their route; Oyster cards making more than 10 trips account for 63.5% of all trips in the sample.

This finding implies that the dataset could be modelled as panel data. This is where the data is multi-dimensional; there are multiple observations made for the same individual. This will be tested in the next section. Further investigation reveals that there are no instances of an individual travelling on more than one OD pair. However 75 Oyster cards are recorded using both modes at least once on their OD pair – indicating that the data is recording behaviour change, which it is hoped will be captured in the next section where choice models are estimated.

## 5.4 Choice Modelling

The models in this section were estimated using Biogeme (Bierlaire, 2003). Two choice model specifications introduced earlier in this chapter were utilised: multinomial logit and cross-nested logit. Each model specification is treated separately in turn, with the introduction, results and discussion completed on the MNL model in advance of the CNL model.

### Model 1a: Mean-Variance Multinomial Logit

The first logit model was estimated based upon the following EU function:

$$EU_{n,OD,m} = \beta_1\mu_{OD,m} + \beta_2\sigma_{OD,m} + \varepsilon_n \quad (5.6)$$

Where  $EU_{n,OD,m}$  is the expected utility for individual  $n$  based on their OD pair ( $OD$ ) and mode choice ( $m$ ).  $\mu_{OD,m}$  and  $\sigma_{OD,m}$  represent the mean and standard deviation of the chosen OD pair and mode combination.  $\beta_1$  represents the marginal utility of mean travel time. Similarly,  $\beta_2$  represents the marginal utility of travel time variation. The error term,  $\varepsilon$ , varies between individuals, denoted by subscript  $n$ .

The alternative hypothesis was that average values of  $\beta_1$  and  $\beta_2$  would be less than zero. Such a restriction was not placed on the ASCs and therefore a 2-tailed t-test was utilised for these parameter estimates. Only the MV variables were included in this first model.

The MNL models were run using the 4,140 records of the non-dominated choice dataset described in Section 5.3.2. The ASCs were estimated in all cases except for Metro (London Underground) which was fixed to zero as the reference case (all mode ASCs would therefore be additive to LU). For each choice situation, two modes were available; the remaining modes would be marked as not available in Biogeme. The results are shown in Table 5.5 below.

**Table 5.5** - Parameter estimates from MNL

<b>Name</b>	<b>Value</b>	<b>t-test</b>
ASC_Metro	0	
ASC_Metro/Light Rail	-0.879	-14.99
ASC_Metro/Heavy Rail	-1.440	-17.34
ASC_Heavy Rail	-2.310	-21.48
B_Standard Deviation	-0.157	-6.92
B_Mean Travel Time	-0.179	-14.63

Adj R<sup>2</sup> = 0.281, Final LL = -2057.95

The first model estimate based upon the Oyster data provides encouraging results. All parameters are significant, and furthermore the parameters associated with the standard deviation and mean travel time are both negative and significant. The null hypotheses on these parameters can therefore be rejected. This model estimates a Reliability Ratio of 0.88. This is within the reasonable range of RR values quoted in the literature, and not too far from the average RP estimate of 1.10 calculated in the literature review of this thesis.

#### **Model 1b: Mean-Variance Multinomial Logit with Interaction Terms**

The result of the estimation of Model 1a showed strongly significant ASCs (i.e. differences between modes). The MNL was therefore estimated again with the addition of interaction terms between mode and standard deviation, as well as between mode and mean travel time. This would capture some of the difference between modes implied by the highly significant ASCs and also allow a RR to be calculated for each of the four modes. As previously noted, the Metro mode was treated as the reference case, and the marginal utility of travel time and standard deviation of travel to would be expected to be less than zero in all cases.



**Table 5.6** - Parameter estimates from MNL with interaction terms

Name	Value	t-test
ASC_Metro	0	
ASC_Metro/Light Rail	0.846	2.85
ASC_Metro/Heavy Rail	-5.05	-9.08
ASC_Heavy Rail	-5.13	-9.11
B_St Dev	-0.093	-3.26
B_St Dev_Metro/Light Rail	-0.304	-7.14
B_St Dev_Metro/Heavy Rail	0.441	5.78
B_St Dev_Heavy Rail	0.0425	0.45
B_Travel Time	-0.173	-9.13
B_Travel Time_Metro/Light Rail	-0.00676	-1.14
B_Travel Time_Metro/Heavy Rail	0.061	3.5
B_Travel Time_Heavy Rail	0.119	7.78

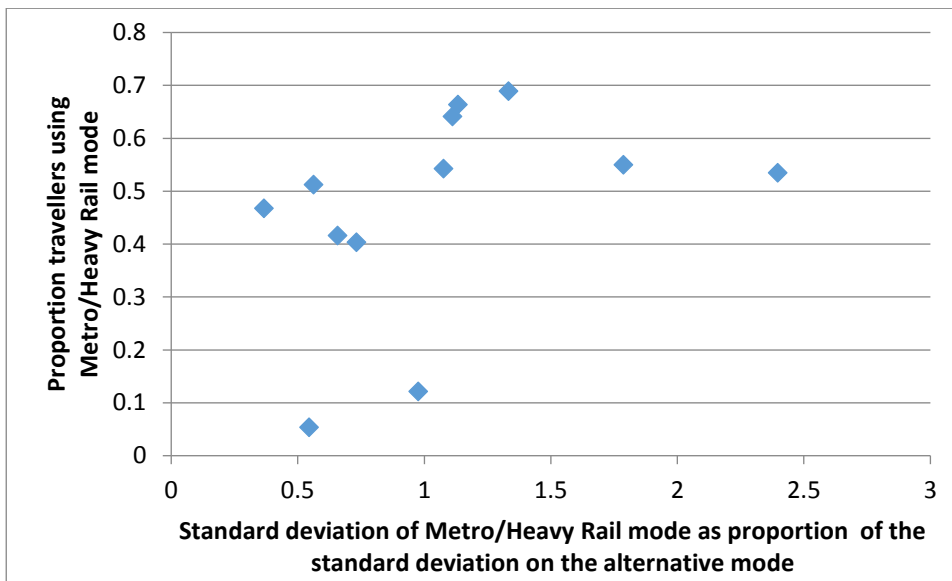
Adj R<sup>2</sup> = 0.312, Final LL = -1964.70

The parameter estimates shown in Table 5.6 present some issues. Although most parameter estimates are statistically significant, there are notable exceptions in the risk parameter for the Heavy Rail mode and the travel time parameter for the Metro/Light Rail mode. Table 5.6 shows the marginal utilities of standard deviation and travel time for each mode excluding such insignificant parameter estimates. As a result of the estimates in Table 5.7 a mode specific RR is also calculated.

**Table 5.7** - Mean-Variance marginal utilities and RR by mode

	B_Standard Deviation	B_Travel Time	RR
Metro	-0.093	-0.173	0.54
Metro/Light Rail	-0.397	-0.173	2.29
Metro/Heavy Rail	0.348	-0.112	-3.11
Heavy Rail	-0.093	-0.054	1.72

Table 5.7 shows that seven of eight parameter estimates are negative as expected. This allows RRs to be calculated for three of the four modes modelled. These range between 0.54 and 2.29, where the latter is substantial but not outside other evidence presented in the literature. The positive risk parameter on the Metro/Heavy Rail mode is against expectations. When this was investigated in greater detail, the data showed that travellers were more likely to use this mode when the standard deviation was higher than for other modes. This is shown in Figure 5.6.



**Figure 5.6** - OD plot of the standard deviation of the Metro/Heavy Rail alternative as proportion of the standard deviation on the alternative mode, against the proportion using the Metro/Heavy Rail mode

This finding may be a fair representation of traveller preferences (i.e. those using Metro/Heavy Rail are risk prone), however a more likely explanation is that at this level of aggregation is that there are too few OD pairs for each mode to accurately represent choice behaviour. This suggests that further estimates of the RR should be made across modes similar to the results shown in Table 5.5 rather than disaggregating MV variables by mode.

**Model 1c: Mean-Variance Multinomial Logit with Additional Variables**

The final MNL model to be estimated includes variables for transfer and headway, but without interaction between mode and MV parameters. The results of this model estimation are shown in Table 5.8.

**Table 5.8** - Parameter estimates from MNL with additional variables

Name	Value	t-test
ASC_Metro	0.000	
ASC_Metro/Light Rail	-0.760	-11.86
ASC_Metro/Heavy Rail	-1.350	-11.58
ASC_Heavy Rail	-2.170	-10.81
B_St Dev	-0.235	-8.82
B_Travel Time	-0.226	-13.90
B_Transfer	-1.300	-5.91
B_Headway	-0.025	-1.80

Adj  $R^2 = 0.288$ , Final LL = -2035.42

The addition of the transfer and headway parameters provides a slightly better fit to the data than Model 1a, but this is not unexpected. Using a single tail test, the transfer and headway parameter estimates are both statistically significant at 5%. This is interpreted as evidence that all variables affect traveller choices on London's transport network. The RR estimated by this model is 1.04 which is reasonable in the context of the literature. Given the improved model fit over Model 1a, as well as an acceptable RR based on parameter estimates that are all negative as expected, this model is the preferred of the three estimated so far in this chapter. A further version of 1c acknowledging the panel nature of the data was run. This was done by allocating a random variable, drawn from a Normal distribution in order to induce correlation for the choice observations of the same Oyster card. The results from this model run are presented in Table 5.9.

**Table 5.9** - Parameter estimates from MNL with additional variables and Panel effects

Name	Value	t-test
ASC_Metro	0.000	
ASC_Metro/Light Rail	-0.760	-12.22
ASC_Metro/Heavy Rail	-1.350	-13.41
ASC_Heavy Rail	-2.170	-10.88
B_St Dev	-0.235	-13.33
B_Travel Time	-0.226	-14.41
B_Transfer	-1.300	-6.67
B_Headway	-0.025	-1.80

Adj  $R^2 = 0.288$ , Final LL = -2035.42

Table 5.9 shows that the inclusion of fixed panel effects across survey subjects does in general increase t-statistics of parameters, but does not have an effect on the model estimation otherwise insofar as the parameter estimates and  $R^2$  / Final LL remain the same as those estimated in the previous model run.

A drawback of the models estimated to this point is their inability to accurately represent correlation between the error terms of similar alternatives. This is addressed by Model 2, using the CNL model specification.

### **Model 2: Mean-Variance Cross-nested Logit**

Model 2 utilized a CNL structure as indicated in Figure 5.1. The  $\alpha$  term defined in Eq. 5 (the allocation parameter) was allowed to vary freely, and each hybrid mode could belong to two nests; that is, Metro/Light Rail could belong to both “Metro” and “Light Rail” nests. Based on experience with the MNL model, interaction terms between mode and travel time/standard deviation were omitted. However ASCs and  $\beta$ s were specified in the same manner. The result of this model estimation is shown in Table 5.10.

**Table 5.10 - Model parameters from CNL**

<b>Name</b>	<b>Value</b>	<b>t-test</b>
ASC_Metro	0.000	
ASC_Metro/Light Rail	-0.845	-2.52
ASC_Metro/Heavy Rail	-1.530	-2.90
ASC_Heavy Rail	-3.340	-14.31
B_Standard Deviation	-0.100	-1.95
B_Mean Travel Time	-0.248	-2.87

Adj  $R^2 = 0.289$ , Final LL = -2028.26

The CNL specification in Model 2 results in negative MV parameter estimates, which suggests a rejection of the null hypothesis; the t-statistic of 1.95 on the risk parameter is significant at 5% (1-tailed). The RR estimate of this model is 0.43. This value is at the lower end of estimates in the literature, but not unreasonable. The ASCs are all significant, indicating differences between the modes. The  $R^2$  and final log-likelihood measures of model fit are lower than the equivalent multinomial logit model, indicating a poorer fit to the data. The allocation and nesting parameters for the CNL

model (i.e. reflecting the degree of membership of a given alternative to each of its member nests) are presented in Table 5.11.

**Table 5.11** - Nesting and Allocation parameters from CNL model estimation

<b>Name</b>	<b>Value</b>	<b>t-test 0</b>	<b>t-test 1</b>
Light Rail nest	1.000		
Metro nest	0.733	2.36	-0.86
Heavy Rail nest	0.220	2.97	-10.52
Metro/Light Rail allocation to Light rail nest	0.755	1.44	-0.47
Metro allocation to Metro nest	1.000		
Metro/Light Rail allocation to Metro nest	0.245	0.47	-1.44
Metro/Heavy Rail allocation to Metro nest	0.855	1.63	-0.28
Metro/Heavy Rail allocation to Heavy Rail nest	0.145	0.28	-1.63
Heavy Rail allocation to Heavy Rail Nest	1.000		

For the nesting structure of the model to be supported, the nesting parameters (the first three rows of Table 5.11) should take a value between one and ten. The light rail nesting parameter is one, as this only has one alternative associated with it. The other nesting parameters are less than one, which suggests that the nesting structure of the model is inappropriate, and therefore conclusions to be drawn from the parameter estimates may be misleading. Furthermore, the allocation parameters are not significantly different from zero and one, which is further evidence against use of the CNL model.

The conclusion must be made that whilst use of the CNL model specification is intuitively reasonable given that there is likely to be correlation between the error terms of the choice alternatives, its use cannot be justified based on the data and results provided in Table 5.11.

## **5.5 Discussion**

The objective of this chapter was to develop a methodology for the estimation of the Reliability Ratio (RR) using public transport smartcard data only. The simple multinomial logit model (1a) achieved this aim, estimating a RR of 0.88. Model 1c (incorporating additional choice variables) was deemed to be an improvement and

estimated an RR of 1.04. Referring to the meta-analysis in Chapter 2 where a mean RP-based RR of 1.10 was calculated, as well as the meta-analysis of Carrion and Levinson (2012) who calculated a mean RR of 1.09, these results appear plausible. The methodology used in relation to the data is fully replicable and it is anticipated that this result will provide the basis for further work in this field. A key issue with the estimates of 0.88 and 1.04 is that the model specification used did not allow for violations of IID.

The CNL model was an attempt to overcome this issue, but had issues itself. Whilst a reasonable RR was estimated, the nesting parameters were not within the expected range, suggesting that the specification of the model could not be justified empirically. It is suspected that there were too few OD pairs within the dataset used and therefore the conclusion is not that this model specification should be rejected out of hand, but rather to recommend that it is estimated using larger datasets containing more OD pairs in the future.

A similar recommendation is made regarding the RR estimates by mode, which was the second of the multinomial logit models. This model estimate was successful insofar as it estimated mode specific RRs for three of the four modes in the model. An unexpected positive parameter value could be observed in the data, but it was not clear whether this was again due to a low sample size of OD pairs for a single mode. Nevertheless, the three positive RRs estimated ranged from 0.54 to 2.29 which are not outside the acceptable range exhibited by the literature.

These results are encouraging; however there are possible developments that could be made based upon the dataset obtained. One criticism of the above method is that it assumes uniform conditions across the entire (3 hour) AM peak dataset. Other studies have favoured percentile based values to populate a standard MV model – which has not been done here.

Finally, the RRs estimated within this chapter should themselves be treated with a degree of caution given data issues identified in Chapter 3. For example, the smartcard records contained times between entry and exit points of a station, and so the parameter estimates are based upon all activities conducted ‘behind the gate’ including platform access time, waiting time etc. in addition to in-vehicle time. The literature tends to focus on in-vehicle time only and therefore the RRs estimated in this chapter are not entirely

comparable. Nevertheless, by taking into account a wider range of elements of the trip, the method arguably provides a closer representation of actual decisions made by travellers.

## **5.6 Conclusion**

This chapter has demonstrated a method of estimating a Reliability Ratio (RR) using RP data. The benefits of undertaking this research are clear; both in terms of the limitations of the alternative SP method (general suitability, high survey cost) and the positive aspects of RP (realism, low survey cost). Combining standard economic and choice modelling approaches formed the basis for estimating the RR from automatically collected data only. Using public transport smartcard data it was possible to estimate a realistic RR during the AM peak period. A number of estimates were made, the majority of which were within an acceptable range when compared to the literature. However use of the more complex but intuitively reasonable model specification, CNL, could not be supported based upon model estimation in this chapter and is rejected, subject to further work.

A range of RR estimates in the literature highlight the importance of estimating a RR which is relevant to its application, an issue which the methodology outlined in this chapter could provide a solution. The objectives related to this chapter are shown in Table 5.12.

**Table 5.12 - Objectives met in Chapter 5**

Objective	Description	Addressed in:
O1	To apply smartcard datasets to a number of real world situations, drawing and developing upon existing studies.	This chapter is a substantive contribution to objective 1 by developing a novel use of smartcard data. The method is applied empirically.
O2	To provide a critique of the smartcard data available (and smartcard data more broadly).	Although the chapter had some success in estimating RRs, some issues with the data were identified. For example, the large dataset became limited in size after choice situations were identified.
O4	To develop a methodology for estimating a VOR using smartcard data.	Despite some not-insubstantial limitations, the methodology developed and applied produced a RR and could form the basis for further work in this area.



Despite efforts to include additional variables in Model 1c, there remains the possibility that there are omitted variables from our model which bias the results. For example, station size or accessibility to platforms may impact upon travellers' mode choice, but this was not explicitly represented in the model due to data availability. The relatively small sample sizes at the OD level of aggregation meant that it was mainly large stations that featured within the dataset, and a station size variable could not be easily modelled. Finally, although it remains likely that the CNL model specification is the most appropriate for representing the choice situation, evidence supporting this claim could not be found in this analysis. This should also be a subject for further investigation.

The assumption of perfect knowledge on the part of every passenger is also unrealistic. A development of the RP method would take into account variability in experience of passengers on a route – this could be performed more accurately where the dataset covered a longer temporal duration than one month. In the discussion, the reader is also advised to be aware of the data description in the previous chapter, and the issues identified with the data. Future research might consider datasets that contain a cost variable so that monetary values of reliability can be estimated directly, thus overcoming the issue of fungibility when utilising a RR alongside separately estimated values of time (Orr et al, 2012).

## **Chapter 6 – Alternatives to Mean-Variance**

### **6.1 Introduction**

This chapter will investigate the underlying utility functions driving traveller behaviour, using the smartcard data. This will involve two elements of investigation.

The first of these elements will return to the alternative indicators of risk identified in Chapter 3, which considered lateness but not earliness. If one of these indicators were to replace the variance or standard deviation in MV, then a different utility function would result. The chapter will initially identify the form that these utility functions would take, and then go on to utilise the methodology developed in Chapter 5 to test whether these indicators provide an improved account of traveller behaviour over the standard deviation.

The second element of this chapter will reintroduce the Scheduling approach to reliability identified and discussed in Chapter 2. The Scheduling utility function, simplified to consist of two straight lines, has been shown in the literature to be capable of representing travel time risk. The functional form does however imply different behaviour to the curved function underpinning MV. This comparison is made in light of the separate literature identified in Chapter 2 which treats MV and Scheduling as equivalent. This will be investigated by firstly matching MV and Scheduling utility functions based upon their properties; and secondly by using the smartcard data to calculate and compare the reliability premia of the functions. This will allow differences between these functional forms to be identified.

### **6.2 Alternatives to Variance in MV**

In Chapter 8 of ‘Portfolio Selection’ (Markowitz, 1959), a range of candidate risk measures were considered as possible replacements for the variance or standard deviation in the MV expected utility function. This was motivated by the observation made by Markowitz that:

*“Analysis based on S [the semi-variance] tend to produce better portfolios than those based on V [the variance]”*

The reason for this being that

*“Variance considers extremely high and extremely low returns as equally undesirable”*

Markowitz, 1959, pp.194

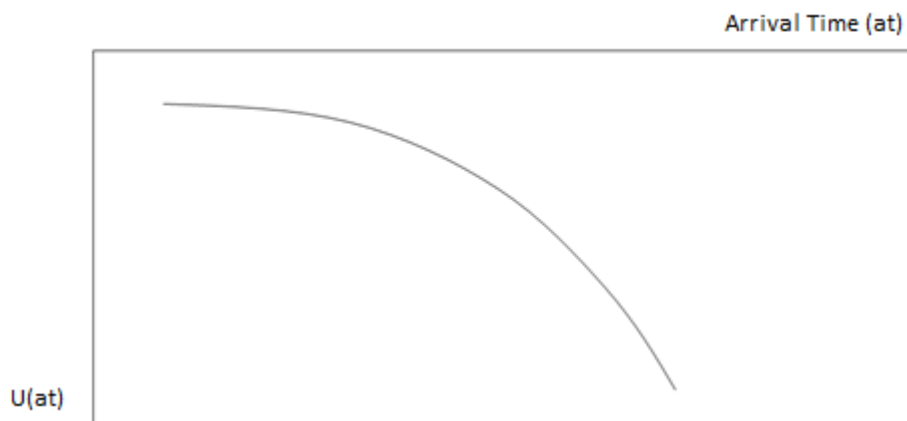
That said, Markowitz did not reject the variance as an indicator of risk. In the same chapter, Markowitz developed this thinking by considering some alternatives to variance, focussed around the idea of financial losses. The justification for this was that risk in losses weigh more heavily on investors than risk in gains. A corresponding narrative in transport could be that travellers are more greatly motivated in their travel decisions by fear of lateness rather than earliness, to the extent that risk in earliness could be omitted from MV analysis.

This links with the outcome of the investigation into alternative reliability indicators conducted in Chapter 3 of this thesis. To re-cap, that analysis identified three indicators, in addition to the standard deviation, worthy of further investigation. These were:

- **Buffer Time** – The difference between 95<sup>th</sup> percentile and median travel times. This would indicate the extra travel time above the median travel time which a passenger should leave to arrive on time in 95% of cases. It could be scaled by dividing it by the median, but it would then lose its simple interpretation.
- **Late trip** – Calculates an average travel time of the 20% longest travel times for a given time period and/or time of day. This indicator could also be scaled by a central measure of travel time.
- **Semi-Variance** – This is the variance calculated for all travellers whose travel time was in excess of the mean travel time. Taking the square root of the semi-variance results in the semi-deviation.

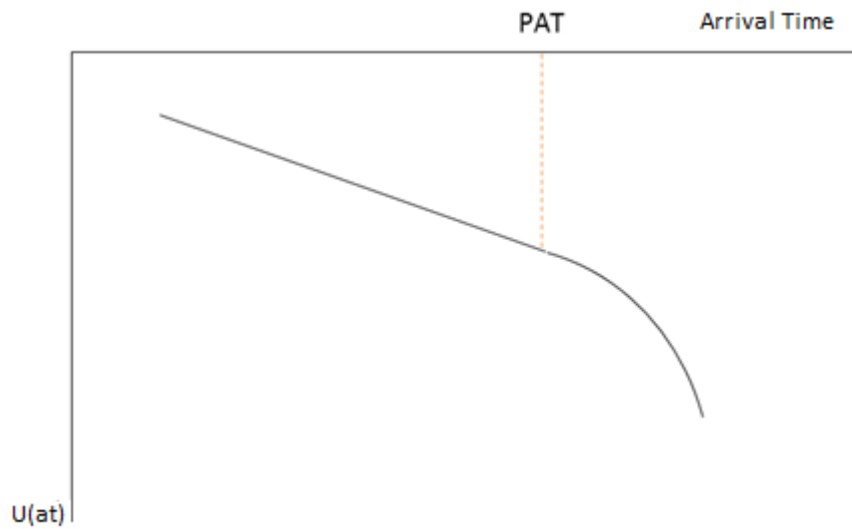
The common element of the above three indicators is their focus on lateness. Like Markowitz, it will be useful to initially consider how these indicators would affect the travellers' utility function.

Chapter 2 established that the utility function for the risk averse traveller would resemble that in Figure 6.1: concave and (usually) monotonically decreasing. This would suggest that risk aversion is evident for all values of  $t$ , as illustrated by curvature on all parts of the function. This risk associated with lateness (higher values of  $t$ ) is greater than that for lower values of  $t$ .



**Figure 6.1** - An example MV utility function for a fixed departure time

The three alternative indicators suggest that only risk in lateness (after the PAT) is of relevance. This would therefore imply risk neutrality (no curvature of the utility function) before the PAT. In what follows, the assumption is made that the PAT exists at some point between the minimum and maximum values for  $t$ . The resulting utility function would therefore resemble that represented in Figure 6.2. It is also possible that the section of the utility function after the PAT could be a straight line, but this will be dealt with in the second section of the present chapter.



**Figure 6.2** – MV utility function (travel time risk after PAT only), for a fixed departure time

Figure 6.2 shows that travel time risk only occurs when there is a possibility of lateness. That is, if earliness could be guaranteed, no risk is considered.

The question naturally arises as to which of Figures 6.1 and 6.2 is the closest account of actual traveller behaviour. To answer this question, the methodology and dataset developed in Chapter 5 will be revisited. Model 1c from that chapter, a MV multinomial logit model containing additional variables for headway and transfers, will be run with each of the four candidate risk indicators in turn. As the dataset is a record of revealed traveller behaviour, the model with the best fit (highest adjusted  $R^2$ , highest final LL) should offer the closest representation of reality. In other words, the risk indicator within the model with the best fit should give the best representation of traveller attitude toward risk in the dataset. Markowitz (1959) showed that standard deviation was associated with the utility function in Figure 6.1, whilst the other indicators more closely represented the utility function in Figure 6.2 as they are measures of risk in lateness only. Lateness is defined as a time after the mean travel time; the PAT therefore is assumed to be equal to the mean travel time. Although this is unrealistic, this simplification allows the model to be estimated without directly asking travellers themselves. The parameter estimates and model fit statistics are shown in Table 6.1.

**Table 6.1** - Model results for each risk indicator using RP dataset

Name	Standard Deviation		Buffer Time		Late Trip		Semi-Deviation	
	Value	t-test	Value	t-test	Value	t-test	Value	t-test
ASC_LU	0.00		0.00		0.00		0.00	
ASC_LUDLR	-0.76	-11.86	-0.45	-5.38	-0.89	-13.26	-0.82	-12.18
ASC_LURAIL	-1.35	-11.58	-0.71	-4.27	-1.48	-12.48	-1.47	-12.34
ASC_RAIL	-2.17	-10.81	-2.30	-11.67	-2.53	-12.49	-2.54	-13.10
B_HEADWAY	-0.03	-1.80	0.01	0.72	0.02	1.53	0.01	0.90
B_RISK	-0.24	-8.82	0.05	5.74	-0.13	-8.03	-0.07	-3.55
B_TIME	-0.23	-13.90	-0.13	-9.95	-0.05	-2.97	-0.17	-11.68
B_TRANSFER	-1.30	-5.91	-0.88	-3.75	-0.73	-3.24	-0.65	-2.75
<b>Reliability Ratio</b>	1.04		-0.36		2.43		0.44	
<b>Final LL</b>	-2035.42		-2057.79		-2040.91		-2067.99	
<b>Adjusted R<sup>2</sup></b>	0.288		0.280		0.286		0.277	

Table 6.1 shows that the Mean-Variance model provides the best fit to the data. This is modelled using the standard deviation in place of the variance as the indicator of risk, in line with convention. All indicators of risk and travel time were expected to be negative.

Table 6.1 suggests that the standard deviation is the best indicator of reliability of those modelled for representing actual traveller choices. It should be noted that the Late Trip model gives fit only slightly worse than the mean-standard deviation. The fit and explanatory power of the semi-deviation model is the poorest, which is unexpected given its similarity to the mean-standard deviation, however the simplifying assumption that the PAT was equal to the mean travel time may have influenced this result. Although the Reliability Ratios calculated are not directly comparable, given the difference in scale of the indicators, the negative Reliability Ratio associated with the Buffer Time is unexpected. This is a result of a positive parameter estimate for the Buffer Time indicator, which would be expected to be negative. As a result, this indicator is rejected.

The finding of Table 6.1 is broadly in favour of the status quo for modelling reliability: using the standard deviation as the indicator for risk. This result supports the type of utility function shown in Figure 6.1, and suggests that utility functions of the type shown in Figure 6.2 are less appropriate. This implies either or both of the following are true:

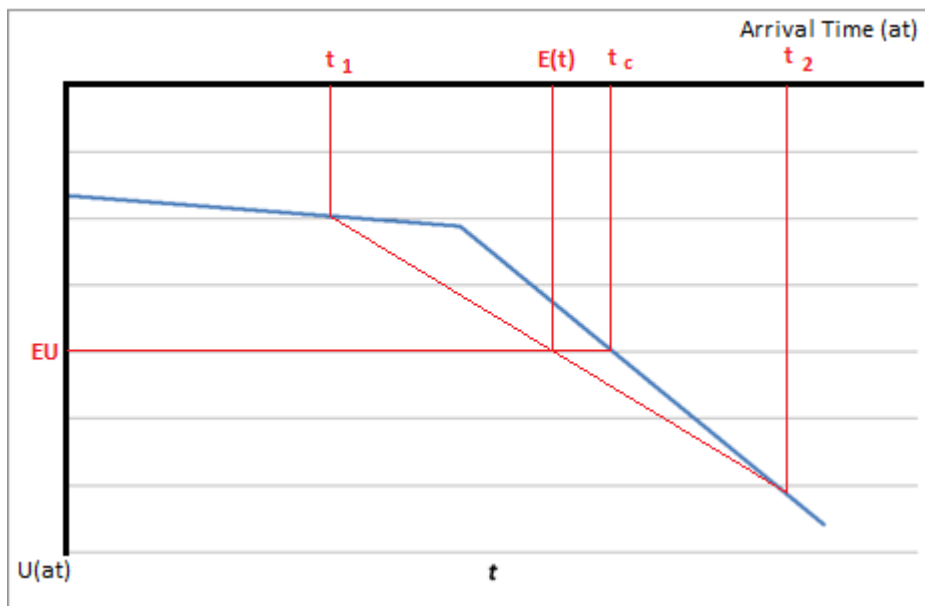
1. Risk in earliness is of value to travellers
2. Linear sections of the utility function are unrealistic

These findings also relate to the Scheduling framework. The Scheduling framework explicitly takes account of travellers' attitude towards earliness, however it also comprises of two (or more) linear sections. Although the Scheduling model cannot be estimated with the RP dataset (as no information regarding traveller PATs is revealed in the RP dataset), it will be useful to understand the characteristics of its utility function that underpins its expected utility. Furthermore, in the light of the work that has established equivalence between Scheduling and MV expected utility functions (Fosgerau and Karlström, 2010), it has been suggested that their respective utility

functions (i.e. underpinning the EU function) should also be approximately equivalent (Batley, 2007). This will be investigated in the next section.

### 6.3 Mean-Variance and Scheduling Utility Functions

To begin a comparison between Scheduling and MV utility functions, it will be useful to visualise an example of the Scheduling utility function to complement the example MV utility function plotted in Figure 6.1. To do this, it is possible to reinterpret Scheduling cost functions (of Small, 1982) into the utility domain, where travel time is combined with *SDE* and *SDL* (see Batley, 2007), and furthermore  $\theta$  takes a value of zero (to introduce a degree of simplification). For typical parameters of a Scheduling EU function where  $\beta > \alpha > \gamma$ , the function would approximately resemble that shown in Figure 6.3. This is expanded with the example of two risky outcomes (similar to that already shown in Figure 2.6) to illustrate the concept of reliability. These two risky outcomes (shown in red) result in an Expected Utility, plotted on the Y axis. Fosgerau and Karlström (2010) establish the equivalence between the optimum Scheduling EU (with respect to departure time) and MV.



**Figure 6.3** - Plot of Scheduling utility function based upon typical parameters where  $\beta > \alpha > \gamma$  (estimated by Li et al, 2010) and omitting a fixed penalty for lateness ( $\theta = 0$ )



Research has recognised similarity between the underlying Scheduling and MV utility functions, with the angle formed between the earliness and lateness sections of the Scheduling function representing an approximation to the risk attitude of the traveller (Batley, 2007). However, the present focus on the *shape* of the utility function has not hitherto been addressed. Moreover, if Scheduling and MV were *formally* equivalent, then intuition suggests that this would require three conditions to hold:

- **Condition I:** Equivalence in the expectations of the utility functions (i.e. as developed by Fosgerau and Karlström, 2010).
- **Condition II:** Equivalence in the average slopes of the utility functions (recognising that slope varies at different points on the functions), where slope is determined by the marginal utility of  $t$  ( $\frac{\delta U}{\delta t}$ ).
- **Condition III:** Equivalence in the curvature of the functions (i.e. reflecting risk attitude), or the changing marginal utility over  $t$  ( $\frac{\delta^2 U}{\delta t^2}$ ).

These conditions will be discussed in greater detail, before being used to combine the Scheduling and MV utility functions.

#### **Condition I:** Equivalence in expected utility

The theory of reliability equivalence discussed in Chapter 2 was focussed upon equivalence in expected utility only. This imposes no substantive restrictions upon the underpinning utility functions; it is feasible that two completely different utility functions could give rise to the same expected utility. This suggests that, for formal equivalence between Scheduling and MV, further restrictions are required so that their respective functions entail similar behaviours on the part of travellers, as follows.

#### **Condition II:** Equivalence in the marginal utility of $t$

The second restriction on equivalence is that the utility functions should imply that travellers have the same overall marginal utility of travel time. The additional restriction to reliability equivalence proposed in this section is that the slopes of ‘equivalent’ functions will be the same, on average. What this restriction effectively implies is that,

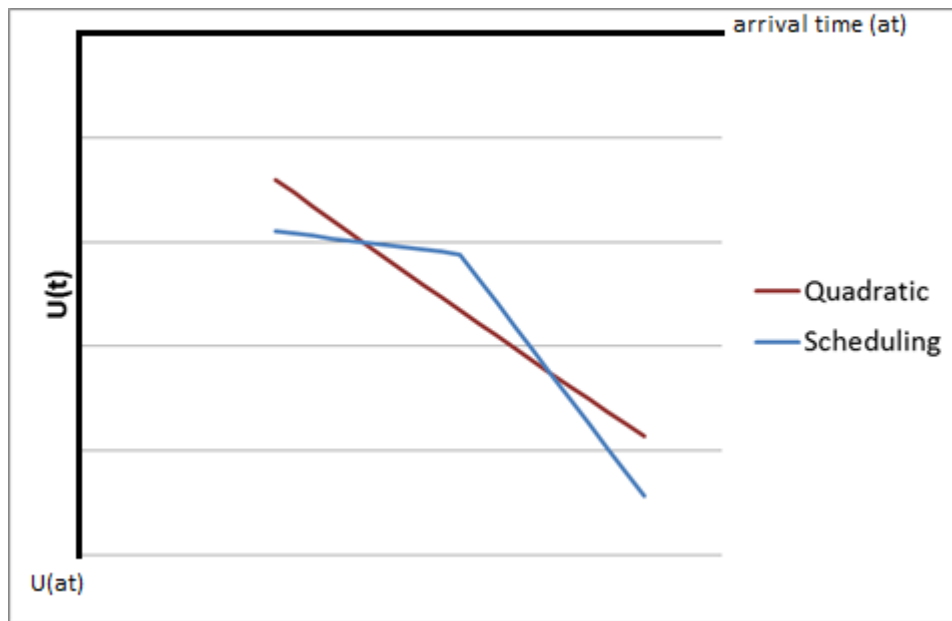
if the effect of travel time risk were removed, then the utility functions would be identical. However, travel time risk is central to the present discussion, and it is therefore useful to understand how well reliability equivalence performs using conditions I and II.

These conditions are applied to fit a quadratic representing MV to the Scheduling function of Figure 6.3; the marginal utility of  $t$  on the Scheduling function was calculated as a weighted average of each of the two slopes. That is to say that, using Pythagoras, the marginal utility of  $t$  for a quadratic ( $U'(t)$ ) is given by:

$$U'(t) = \left( (\alpha - \beta) \frac{L(Early)}{L(Early) + L(Late)} \right) + \left( (\alpha + \gamma) \frac{L(Late)}{L(Early) + L(Late)} \right) \quad (6.1)$$

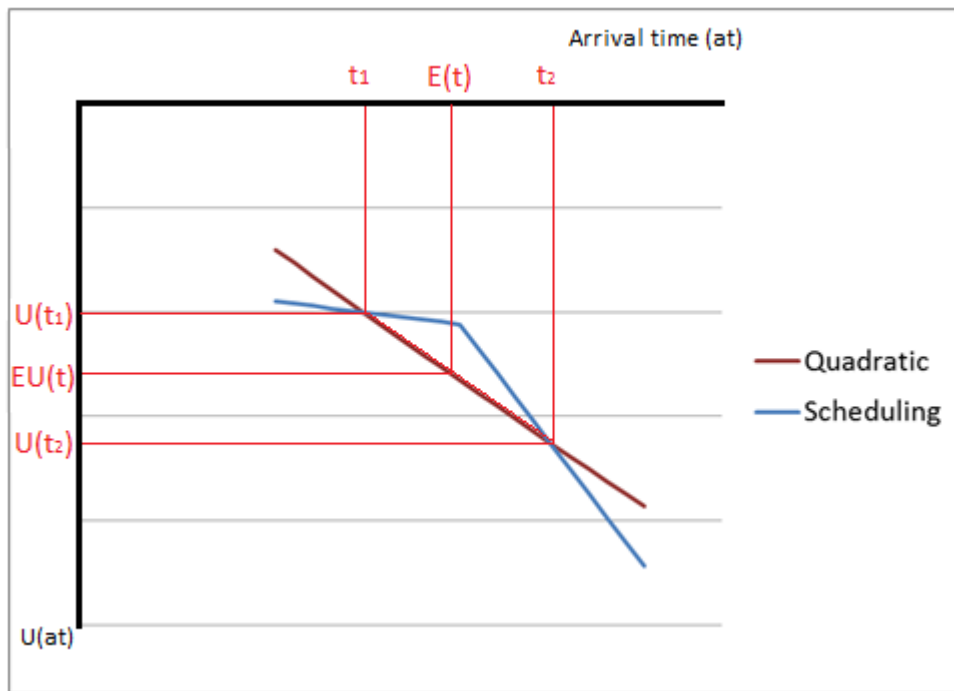
Where  $L(Early)$  is the length of the Scheduling utility function prior to the PAT, and  $L(Late)$  is the length of the Scheduling utility function after the PAT.

The result of this process is that the two functions now somewhat resemble one another, but substantive differences still remain. An example of such differences is given by Figure 6.4.



**Figure 6.4** - Plot of equivalent utility functions, where a quadratic function is fitted to a Scheduling function based upon overall EU and average slope of the Scheduling function

The quadratic in Figure 6.4 is approximately a straight line. It crosses the scheduling function at two points. If two equi-probable values of  $t$  were defined at these crossing points, then it would be possible to suggest that – according to conditions I and II – the functions were approximately equivalent, as shown in Figure 6.5.



**Figure 6.5** - A two outcome example demonstrating the possibility of equivalence between Scheduling and MV utility functions based upon two conditions of equivalence

Despite this apparent equivalence based upon two conditions, the utility functions remain quite different and would therefore imply differing behaviour on the part of the traveller. This discrepancy is particularly obvious around the Scheduling PAT and at the extremes of the utility functions. This suggests that equivalence in EU and the marginal utility of  $t$  only will not result in comparable utility functions. It is for this reason that the third condition for utility equivalence is now introduced.

**Condition III:** Equivalence in the changing marginal utility over  $t$

The microeconomic literature has demonstrated that it is the curvature of the utility function that represents risk (Arrow, 1970). Polak (1987) interpreted the curvature of the utility function in a travel time context as related to travel time risk. Accordingly, the final condition for equivalence between Scheduling and MV utility functions is commonality in the change in marginal utility over travel times – or the curvature of the function. The Scheduling function of Figure 6.3 is comprised of two linear sections. However, the angle made between these two linear sections,  $\theta$ , implies a risk attitude. This is shown in Figure 6.6.



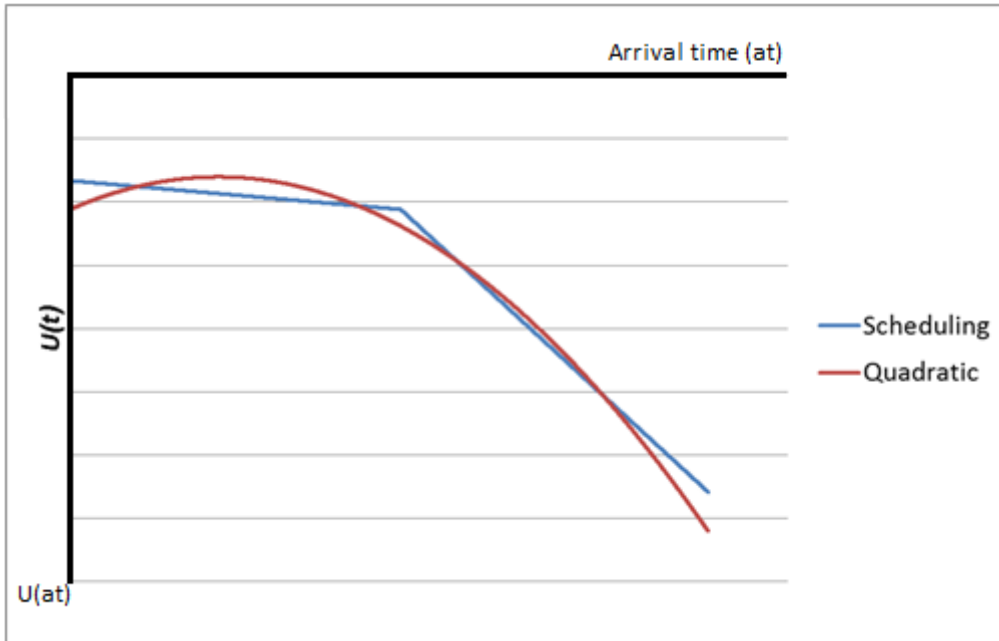
**Figure 6.6** - A Scheduling utility function, with the risk angle,  $\theta$

With reference to Figure 6.6, if  $\theta$  takes a value between  $0^\circ$  and  $180^\circ$ , then the function would imply risk aversion. If the value of  $\theta$  were greater than  $180^\circ$  but less than  $360^\circ$  then risk proneness would be implied.  $\theta$  taking a value of exactly  $180^\circ$  would imply risk neutrality. In the case of the quadratic, the curvature of the function indicates the risk attitude of the traveller. It is therefore postulated that some relationship exists between the *risk angle* of the Scheduling utility function and the second derivative of the MV utility function. This is the basis of the third condition for equivalence between the functions: that the risk attitude implied by the two functions is the same, measured in the terms set out in this paragraph.

This relationship can be illustrated by using a number of hypothetical Scheduling functions along with conditions I and II. The difference between each Scheduling function and the fitted MV function is minimised subject to the second derivative of the quadratic, which in the case of the standard quadratic  $at^2 + bt + c$  would be  $2a$ . The relationship between  $2a$  and  $\theta$  can therefore be estimated. The hypothetical Scheduling utility functions were formed by allowing each of  $\alpha$ ,  $\beta$ , and  $\gamma$ , to take one of three values, resulting in 27 possible functions. A linear relationship between  $2a$  and  $\theta$  was observed, which was formally estimated as:

$$2a = -0.208 + 0.001 \theta \quad (6.2)$$

Drawing upon all three conditions, it is now possible to plot the ‘equivalent’ MV (quadratic) utility function to the Scheduling function of Figure 6.3.

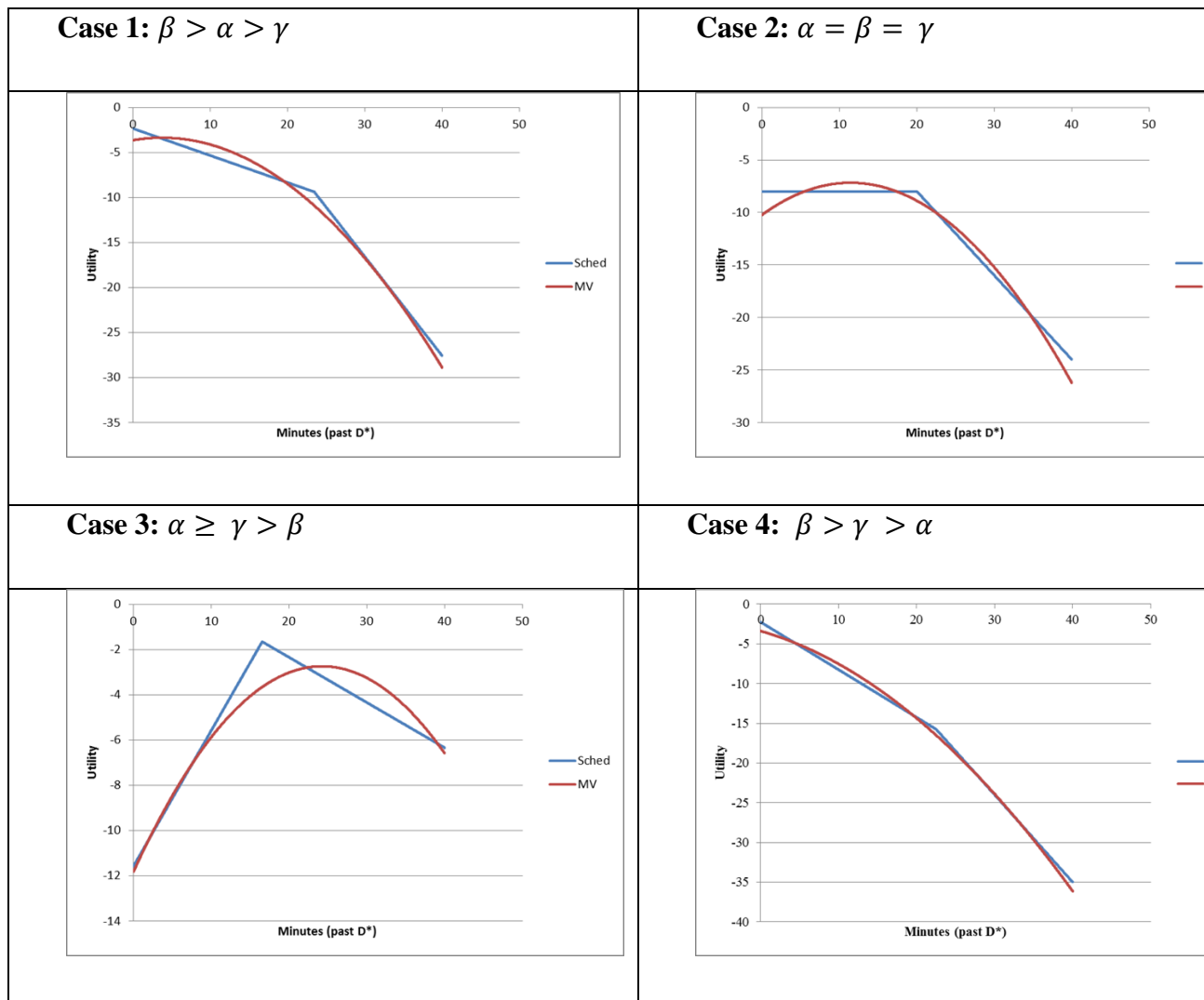


**Figure 6.7** - A quadratic utility function fitted to a simple Scheduling utility function, based upon the three conditions

Based upon the three conditions defined above, the two utility functions now closely resemble one another, implying not only equivalence in EU, but also approximate equivalence in the marginal utility of  $t$  and risk attitude.

### 6.3.1 Issues With Matching MV and Scheduling Utility Functions

To further illustrate the method outlined above, conditions I, II and III were employed to fit a quadratic to a range of different Scheduling utility functions. Some exemplars are shown in the matrix below:



**Figure 6.8** - A quadratic utility function fitted to four Scheduling utility functions, based upon the three conditions of equivalence

From the fitted utility functions in Figure 6.8, a number of common issues are observed:

1. The four crossing points between the functions result in three parts of the quadratic function where a lower utility is estimated than the corresponding Scheduling function. There are also two sections where the quadratic estimates higher utility for given values of  $t$ .
2. The continuous quadratic function will often display both positive and negative gradients of the utility function, even when both sections of the Scheduling utility function have a negative slope. This has the potential to introduce different risk attitudes between the functions. Focussing on the quadratic of

Figure 6.7, the quadratic has a positive slope for smaller values of  $t$ . This may however have an intuitive interpretation, since where there are very ‘early’ travel time outcomes (in relation to the PAT), travellers may be more likely to favour a risky travel time prospect.

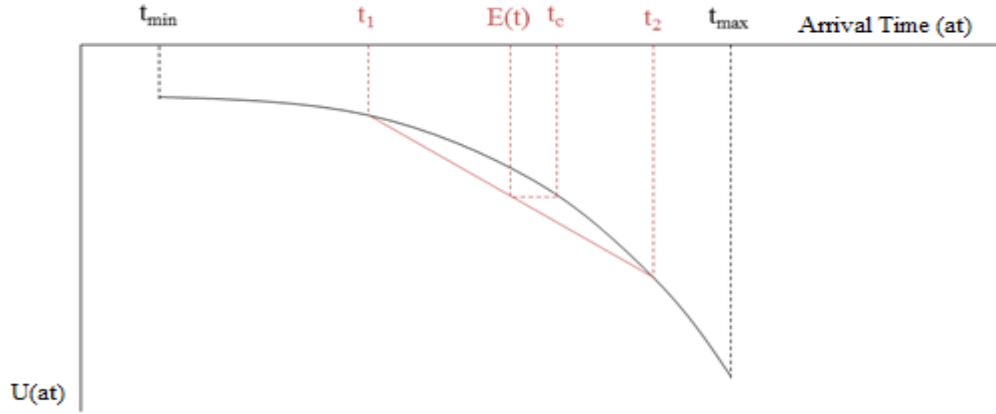
3. The quadratic form struggles to adequately represent the change in gradient of the Scheduling utility function at the PAT. In all examples, a discrepancy between the utility functions was observed in this region. This is of particular concern, since it is this region of the utility function where the impact of reliability issues will be most felt.

It is arguably issue 3 that presents the biggest issue for equivalence between utility functions, particularly in the case of shorter distance urban public transport. This is because these small deviations in arrival time from the PAT are likely to be the most frequently experienced. This may affect VOR estimates. Quantitative analysis of this issue is provided in the next section.

### **6.3.2 The Reliability Premium**

To further illustrate the issue of discrepancies between the utility functions around the PAT, the ‘reliability premium’ was calculated for the utility functions shown in Figure 6.8. The reliability premium was developed by Batley (2007) and is analogous to the risk premium developed by Pratt (1964). It has subsequently been adopted by other researchers (Beaud et al, 2016). This value is the amount that a traveller would pay, in minutes, to avoid a risky situation.





**Figure 6.9** - A concave utility function with two risky travel time outcomes, the expected value of  $t$ ,  $E(t)$ , and the certainty equivalent,  $t_c$

To calculate the reliability premium for a concave function, the expected value of the risky prospect is given by  $E(t)$ . The utility of  $E(t)$  can be calculated using the utilities and probabilities of  $t_1$  and  $t_2$ .

$$UE(t) = U(t_1) + p(t_2)(U(t_2) - U(t_1)) \quad (6.3)$$

$E(t)$  is then subtracted from the certainty equivalent  $t_c$ . Figure 6.9 shows that the utility for  $E(t)$  is the same of that for  $t_c$ . Using the utility function  $f(t)$ , it is straightforward to calculate  $t_c$ .

Batley (2007) showed how to calculate an equivalent value for a Scheduling function. The utility of  $E(t)$  can be calculated as shown in Equation 6.3. The value of  $t_c$  can now be calculated with reference to  $f(t)$ , which in this case was defined by Small (1982). With reference to Figure 6.3 it is possible that the certainty equivalent may fall before or after the PAT. (Batley (2007) also considers the situation where  $t_c$  is equal to the PAT). If  $t_c$  falls before the PAT, then it is given by:

$$t_c = \frac{U(t_c) - \beta PAT}{(\alpha - \beta)} \quad (6.4)$$

Similarly,  $t_c$  after the PAT can be calculated using equation 6.5:

$$t_c = \frac{U(t_c) - \gamma PAT - \theta D_L}{(\alpha + \gamma)} \quad (6.5)$$

The reliability premium is now calculated based upon the four MV and Scheduling utility function comparisons in Figure 6.8. For purposes of illustration, two equiprobable outcomes are defined, which can be either  $\pm 1, 3$  or 5 minutes from the PAT. Reliability premia for the ‘equivalent’ functions of Figure 6.8 are presented below in Table 6.2.

**Table 6.2** - Reliability premia calculated for four cases of matching a MV utility function to a Scheduling utility function. All values presented in units of travel time,  $t$ .

	Case 1		Case 2		Case 3		Case 4	
$t$ (relative to PAT)	Sched	Quad	Sched	Quad	Sched	Quad	Sched	Quad
$\pm 1$ min	0.36	0.03	0.50	0.06	-0.67	-0.07	0.23	0.01
$\pm 3$ mins	1.09	0.23	1.50	0.50	-2.00	-0.63	0.68	0.13
$\pm 5$ mins	1.82	0.62	2.50	1.33	-3.33	-1.64	1.14	0.37

Table 6.2 shows that for a risky prospect closest to the PAT, the differences between the premia calculated for Scheduling and quadratic functions are greatest. The change in gradient of the Scheduling utility function creates a relatively large space between the expected outcome and the utility function. Moreover, the Scheduling utility function has a larger reliability premium than the quadratic at all three levels tested. In proportionate terms, this difference is largest for a prospect where the outcomes are closest to the PAT, and smallest for a prospect where the outcomes are furthest from the PAT.

This difference will have practical implications for the valuation of reliability. If a survey participant responds to reliability situations in line with a quadratic utility function, but the analyst assumes that they are utilising a Scheduling function, then the participant will appear less sensitive to risk than is the case in reality. This shows that despite measures to ensure similarity between Scheduling and MV utility functions (i.e. through conditions I, II and III), there are intrinsic differences between them. This

suggests that MV and Scheduling are not equivalent utility functions, and that they necessarily imply different behavioural traits of travellers.

This finding therefore provides an additional explanation to the findings of work such as Börjesson et al (2012) who could not find empirical evidence of the theoretical equivalence between MV and Scheduling proposed by Fosgerau and Karlström (2010). The work above suggests that the underlying utility functions, which ultimately drive the expected utility, are only approximately equivalent. Moreover, the correspondence between the utility functions is low around the PAT. This suggests that the shape of the Scheduling Utility function is problematic due to the kink at the PAT. The expected utility calculated based upon the assumption of a quadratic utility function will be different to that calculated based upon the Scheduling utility function, even when the three conditions for equivalence are satisfied. This provides one reason, among others in the literature, for the approximate nature of the equivalence established by Fosgerau and Karlström.

## **6.4 Conclusion**

This chapter has focussed on the utility function of travellers, particularly with respect to travel time risk. It has investigated alternatives to the standard curved utility function often assumed to underlie departure time decisions.

In the first instance, alternative indicators of risk, identified by analysis in Chapter 3, were tested using the RP modelling framework and data proposed in Chapter 5. This analysis found that in the context of the London public transport network, the mean-standard deviation expected utility function provided a better account of choice behaviour than indicators based upon some measure of lateness only. This result supports the standard quadratic-type utility function.

The next section of the chapter investigated another alternative utility function – the Scheduling utility function. An attempt was made to fit the standard curved utility function to the Scheduling function based upon the marginal utility of travel time and risk implied by the scheduling function. Despite this fitting process, substantive differences remained between the functions, suggesting that they should not be treated as interchangeable. This was particularly the case for short distance urban travel, as the

differences observed using the reliability premium were especially evident. This would support use of a smooth continuous utility function that underpins the MV framework. The result provides a further explanation for empirical discrepancies between MV and Scheduling expected utilities; that the utility functions underlying each framework are fundamentally different, even when matched based upon the three conditions proposed.

Table 6.3 shows how the present chapter contributes to the thesis objectives.

**Table 6.3** - Objectives met by Chapter 6

Objective	Description	Addressed in:
O1	To apply smartcard datasets to a number of real world situations, drawing and developing upon existing studies.	In Section 6.2, the smartcard data was applied along with the methodology developed in Chapter 5 and indicators calculated in Chapter 3 in order to evaluate which indicator was most appropriate.
O6	To explore improvements to the standard Mean-Variance framework, including other statistical indicators of risk, the shape of the utility function and potential alternative frameworks.	Both Sections 6.2 and 6.3 contribute to this objective: in 6.2 a number of indicators were investigated in relation to the shape of their utility function. Their appropriateness to describe choice behaviour was tested using the RP dataset from Chapter 5. In Section 6.3 a typical curved MV utility function was compared to a scheduling function and differences were identified.

Ultimately, this chapter supports the use of the mean-standard deviation model for the treatment of reliability. The curved, risk averse and monotonic utility function appears to best represent traveller choices on one hand, whilst the linear sections of the utility function implied by alternative indicators and the Scheduling utility functions imply behaviour that is not consistent with this.

## **Chapter 7– Conclusion**

### **7.1 Introduction**

This chapter will attempt to draw the previous six chapters together. The first part of the conclusion provides a summary of what has been done in each of the chapters. The second part of this concluding chapter returns to the aims and objectives identified in Chapter 1. The three aims of this study were:

A1. To develop understanding of public transport smartcard data and identify key strengths and limitations through application of the data;

A2. To apply smartcard datasets to the Mean-Variance framework to improve understanding of reliability and passenger responses to it;

A3. To conduct a comprehensive review of the Mean-Variance framework and investigate possible improvements.

This chapter will assess to what extent these aims have been met. This will be done through identifying whether objectives associated with each of the aims were achieved. The chapter will continue by identifying some of the limitations of the study, which will in turn lead to the identification of further work, and finally some concluding remarks.

### **7.2 Summary of the Thesis**

**Chapter 1** introduced the topics of this thesis which were travel time reliability and smartcard data. In the case of the former, definitions of reliability in transportation were provided, and the primary frameworks were introduced. Smartcard technology was also introduced, with a focus upon public transport applications. This provided sufficient background to lead into a section on the motivation, aims and objectives of the thesis. This was supported by transport policy, both at the international level (UK and the Netherlands) as well as at the level of the city – namely London, from which the data used in this thesis was sourced.

**Chapter 2** was the literature review, which expanded upon the themes introduced in Chapter 1 and provided justification for the aims and objectives of the thesis. The

literature review introduced expected utility theory which would form the theoretical basis for much of the research that followed. A key aspect of Chapter 2 was an introduction to the Mean-Variance framework for the treatment of reliability. The primary use of this framework is the estimation of the value of reliability (VOR). Examples of VOR studies were outlined, specifically in relation to the data collection methodology they employed (namely Stated Preference (SP) and Revealed Preference (RP)). The benefits and drawbacks of each methodology were discussed, and the results compared. An attempt was made to identify the factors influencing the Reliability Ratio (RR), although no clear conclusion could be made. This chapter also outlined theoretical and empirical contributions to what was termed 'Reliability Equivalence', in advance of work that followed in Chapter 6.

**Chapter 3** was focussed upon the smartcard data. The initial part was a literature review of studies that have made use of similar datasets. The datasets used by this study were then described, and potential issues arising from their use were identified. Following some exploratory analysis, candidate statistical indicators of reliability were introduced and compared using the data. Four of these indicators were identified as suitable for further analysis, which also followed in Chapter 6.

**Chapter 4** was the second substantive application of the smartcard data. Building upon similar studies on private and public transport modes, this chapter identified a range of variables that may have had an impact upon reliability levels on the London Underground. Linear regression models were estimated in order to identify the impact of these variables on the standard deviation of travel time. The work was partially successful in identifying the factors affecting reliability, but the models were of poor explanatory power, which may have been due to omitted variables and the definition of reliability used (which was related to the data available).

**Chapter 5** provided a substantive development in relation to the literature. The primary contribution of the chapter was to treat smartcard data as an RP data source and estimate a Reliability Ratio using the MV framework. The chapter described how the dataset was identified, and how it was applied alongside standard choice models. The models estimated Reliability Ratios that were broadly in line with those found in the



literature. The use of a cross-nested logit (CNL) model structure was not supported by the data – this was likely due to the limited size of the dataset.

**Chapter 6** focussed upon the shape of a traveller’s utility function. The first substantive element of the chapter was to identify the shape of the utility function implied by the candidate reliability indicators identified in Chapter 3. These indicators were then each tested using the modelling framework developed in Chapter 5, which found that the standard deviation was the best fit to the data. The second substantive element of the chapter was a comparison of the shapes of MV and Scheduling utility functions. The MV utility function was matched to the Scheduling utility function based upon three conditions. Despite this process, substantive differences between the functions remained, which suggests that the functions imply separate behaviours on the part of the traveller. This provides one account of the approximate nature of reliability equivalence and a potential explanation for discrepancies between frameworks found by empirical research in the field.

### 7.3 Aims and Objectives

The above contributions can be referenced against the aims and objectives set out in Table 1.1 at the outset of this thesis. In the table below (Table 7.1), the relevant objectives are listed under each of the aims. Under each of these is how the objective was met in the thesis.

**Table 7.1** - How the thesis has met the aims and objectives defined at the outset

<b>Aim 1:</b> To develop understanding of public transport smartcard data and identify key strengths and limitations through application of the data;	
<b>Objective 1:</b> To apply smartcard datasets to a number of real world situations, drawing and developing upon existing studies.	<b>Objective 2:</b> To provide a critique of the smartcard data available (and smartcard data more broadly)
Chapter 3 outlined existing smartcard studies and demonstrated how the data	The smartcard literature review in Chapter 3 allowed general issues with the

<p>could be handled and analysed in order to calculate key reliability indicators.</p> <p>Chapter 4 applied the data with linear regression models to understand the factors influencing reliability levels on the London Underground. It showed that a relationship between the dependent and some explanatory variables existed, but the model was not of adequate explanatory power to forecast reliability levels.</p> <p>Chapter 5 applied multi-modal smartcard data alongside existing choice modelling methods to demonstrate that smartcards could be used as an RP data source.</p> <p>Chapter 6 combined candidate reliability indicators from Chapter 3 with the method developed in Chapter 5 to identify the most suitable indicator.</p>	<p>data to be identified. Further issues were identified after initial analysis.</p> <p>Chapters 4 and 5 identified specific issues related to the data through application. These chapters were able to demonstrate useful and novel applications of the data.</p>
<p><b>Aim 2:</b> To apply smartcard datasets to the Mean-Variance framework to improve understanding of reliability and passenger responses to it</p>	
<p><b>Objective 3:</b> To develop the means for improving understanding of the factors affecting Transport Reliability using smartcard data</p>	<p><b>Objective 4:</b> To develop a methodology for estimating a VOR using smartcard data</p>
<p>Chapter 3 demonstrated how these large datasets could be effectively handled and</p>	<p>Chapter 2 provided an introduction to SP and RP in turn, highlighting issues with each data collection methodology.</p>

<p>analysed in order to calculate key variables related to reliability.</p> <p>Chapter 4 drew upon previous studies and experience of Chapter 3 to estimate linear regression models which established those variables having an impact upon reliability levels.</p>	<p>A comparison of the results from the frameworks made the case for further investigation into RP.</p> <p>Chapter 3 demonstrated how the data could be analysed in relation to reliability indicators. It also highlighted the issues related to the data.</p> <p>Despite limitations in the analysis, Chapter 5 demonstrated that a Reliability Ratio could be calculated using the smartcard data.</p>
<p><b>Aim 3:</b> To conduct a comprehensive of the Mean-Variance framework and investigate possible improvements.</p>	
<p><b>Objective 5:</b> To review the origins of the Mean-Variance framework from its origins in finance to its transition and use in transport contexts.</p>	<p><b>Objective 6:</b> To explore improvements to the standard Mean-Variance framework, including other statistical indicators of risk, the shape of the utility function and potential alternative frameworks.</p>
<p>EUT was introduced in Section 2.1. A thorough introduction to MV, its link to EUT and finance was provided in Section 2.2. The transition of MV to transportation contexts was investigated in detail in Section 2.2.2</p>	<p>Research which has examined alternative indicators of travel time risk was discussed in Section 2.2.3 of Chapter 2. The alternative frameworks were introduced in Section 2.3, which included potential equivalence between MV and Scheduling.</p> <p>Section 3.4.2 identified alternative statistical indicators of travel time risk</p>

	<p>for MV. These indicators were then compared using the smartcard data.</p> <p>Both Sections 6.2 and 6.3 contributed to this objective:</p> <ul style="list-style-type: none"> <li>-in Section 6.2 the candidate statistical indicators were investigated in relation to the shape of their utility function.</li> <li>-in Section 6.3 a typical curved MV utility function was compared to a Scheduling function and differences were identified. The discrepancy around the PAT was the greatest, and use of a continuous utility function was identified as preferable.</li> </ul>
--	--

To summarise the table above, it will be useful to describe the key conclusions of the thesis in relation to its stated aims.

**Aim 1:** To develop understanding of public transport smartcard data and identify key strengths and limitations through application of the data.

This aim has been met primarily through Chapters 3 to 5. These Chapters have demonstrated strengths of the data insofar as it has been successfully applied to achieve the following:

- Chapter 3 – The calculation and comparison of reliability indicators
- Chapter 4- The identification of the factors affecting reliability levels
- Chapter 5 – The estimation of Reliability Ratios using smartcard data only

**Aim 2:** To apply smartcard datasets to the Mean-Variance framework to improve understanding of reliability and passenger responses to it.

Chapters 4 and 5 made the most substantive contribution to this aim in the thesis:

Chapter 4 – The identification of factors influencing the standard deviation of travel time.

Chapter 5 – The estimation of parameters related to the marginal utility of travel time and standard deviation of travel time.

**Aim 3:** To conduct a comprehensive of the Mean-Variance framework and investigate possible improvements.

Chapters 2, 3 and 6 were the key chapters contributing to the fulfilment of this aim.

Chapter 2 – A detailed introduction to MV, including the early work in the field of portfolio theory. It highlighted potential improvements to MV through using alternative indicators of risk, as well as introducing the Scheduling and Mean-Lateness frameworks.

Chapter 3 – Analysis conducted on a long list of candidate reliability indicators.

Chapter 6 – Suggested that the standard deviation of travel time was the best indicator of travel time risk. This chapter also investigated the shape of the underlying MV and Scheduling utility functions, and found discrepancies between them.

## **7.4 Limitations of the study**

Some of the limitations of this thesis have been highlighted at relevant stages. Although the thesis encourages use of RP as an alternative to SP, both general and specific issues remain, which are outlined here.

There are some limitations of this thesis related specifically to the smartcard datasets. This is not to say that these smartcard datasets were unsuitable for the application, but is a reflection of a decision to make use of generic smartcard datasets rather than those collated for a specific purpose. For example, in Chapter 5 once Dataset 2 had been cleaned and limited to OD pairs where a non-dominated choice was available to travellers, some modes had very few observations. This meant that a single highly trafficked OD pair could impact upon the estimation of travel time and risk parameters.

In one instance this implied that travellers were risk prone on a mode. Furthermore, the relatively small choice set did not allow more complex model structures such as the CNL to be estimated. Therefore, the preferred model estimates, based upon an MNL structure, were unable to take correlation between error terms of alternatives into account. This would mean that the model would inadequately take patterns of substitution into account. As a result of this shortcoming of the model and data, the recommendation is made that further efforts to estimate choice models using these data are made based upon a larger initial choice set, or a dataset collected for the purposes of estimating an RP model.

Chapter 2 also identified a caveat to the data which impacts upon the Reliability Ratios obtained: all travel times and therefore travel time variation is from station gate to station gate. This means that the travel times recorded cover a number of activities of passengers behind the gate, including waiting time and accessing the platform among others. Reliability Ratios in the literature are usually calculated based upon in vehicle travel time only. Therefore, although the values obtained in Chapter 5 were apparently sensible in relation to the literature, they could not be directly compared to other values. One method of overcoming this issue would be to calculate monetary values of marginal travel time and reliability rather than marginal utilities and apply them only to that context. However the zone-based fare structure does not allow this to be done straightforwardly – the modelling would require additional modes with separate costs for the same trip. This also raises the question of the completeness of the choices in the model when only rail based public transport modes were included – omitting bus, car and active mode-based trips could have biased model results. Although Chapter 5 only found limited evidence of travellers using a bus alternative where rail was available, strictly speaking the bus mode should have been included.

An issue also relevant to the choice modelling is the stated assumption that travellers have perfect information about the travel time distribution of the trip that they are about to undertake, as well as the travel time distribution of the alternatives. Although the assumption has been made clear in this thesis and similar work, it is nevertheless unrealistically strong.

Chapter 4 was only partially successful in achieving its objective; it was able to identify factors affecting reliability, but the explanatory power of the models was poor and it was concluded that they could not be usefully employed to predict future reliability levels. A key reason for this was that the smartcard data was not sufficient for the latter task; actual passenger travel times would vary behind the barrier due to many different circumstances, and little was known about the actual operation of the transport system. A case could be made for estimating these models based upon more traditional data collection methods e.g. observing passengers travelling on the network rather than the one proposed here.

The final substantive chapter of the thesis, Chapter 6 made a useful contribution to the thesis insofar as it questioned the most appropriate form of the utility function to represent travel time risk. The chapter concluded that the shape of the Scheduling utility function was an approximation to the curved MV function, but it did not go so far as to provide firm evidence as to the preferred framework. The interpretation of an observed discrepancy between the functions was that it was one source of error between the Scheduling and MV frameworks. It is concluded that further empirical evidence, backed up by theory, is needed.

### **7.5 Further work arising from this study**

The main aspect of this thesis has been the development of an RP methodology in light of a literature which suggests that such data sources should be preferred to those based on hypothetical questions. Although the literature has to some extent compared RP and SP through meta-analyses, a useful extension to this thesis would be to conduct an SP on London's public transport network and compare the results to the RP method. A similar result should be observed if the experiment were designed correctly, although the literature would seem to suggest that there are intrinsic differences between the two methods of data collection.

Another strand of further work is also related to the limitations of the RP method. An obvious recommendation would be to collect data specifically for the purpose of estimating choice models, so that parameters for travel time and travel time risk could be more accurately estimated as the sample sizes could be greater. Another

recommendation is to extend the scope of such modelling to include the bus mode. Including this mode would introduce a cost variable and result in the choice set being more fully specified. This proposal was given fuller consideration as part of this thesis but was not implemented due to time and data constraints. The method of modelling bus choices would be to link the smartcard data to bus AVL data as suggested by Wei (2010) to obtain a bus OD matrix, similar to the rail OD matrices used in this thesis.

The Oyster sample sizes in Chapter 4 appears sufficient for the task of estimating regression models, yet those estimated had poor explanatory power. This method could therefore be improved with supporting data. These could include actual train performance data to capture in vehicle time. There could also be focus on the non-train-based elements of travel; acknowledging that TTV could be incurred, for example, walking to the platform from the station gate. Such a study would benefit from validation against real observation or pedestrian microsimulation in order to uncover the factors driving reliability levels with greater confidence.

## **7.6 Summary and Contribution**

The origin of this thesis was to better understand potential applications and drawbacks of smartcard data in the context of public transport reliability for the project sponsor, TfL. The literature review undertaken indicated that using actual choice data to estimate valuations of reliability was a worthwhile strand of research.

The RP methodology was developed with the MV framework in mind, and represented a contribution to knowledge through developing a novel method for the estimation of the VOR. Chapter 4 also demonstrated that the data could be applied in order to understand the factors affecting reliability on the London Underground.

The final substantive chapter, Chapter 6, developed upon the analysis of reliability indicators in Chapter 3 and found evidence that the standard deviation should be the preferred indicator of risk in modelling choice behaviour. A comparison of the MV and Scheduling utility functions found that there were differences between them, which could have implications for the VOR.



This thesis has found that smartcard data can make a useful contribution to the analysis of public transport reliability. The thesis has delivered a detailed account of MV and its origins, and has exploited smartcard data to investigate reliability indicators and the factors affecting reliability on the London Underground, and to calculate a Reliability Ratio for rail based modes in London. The latter two tasks were however only partially successful, and the results thus come with some caveats. Nevertheless, the thesis has made material progress in developing understanding of an emerging form of public transport data; an understanding it is hoped can be built upon by subsequent researchers.

## References

- ABKOWITZ, M. D. & ENGELSTEIN, I. 1983. The Factors Affecting Running Time on Transit Routes. *Transportation Research Part A: General*, 17, 107-113
- ABKOWITZ, M. D., SLAVIN, H., WAKSMAN, R., ENGLISHER, L. & WILSON, N. 1978. *Transit service reliability*. Cambridge, MA: USDOT Transportation Systems Center and Multisystems, Inc.
- ALLAIS, M. 1953. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 503-546.
- ARROW, K.J. 1974. *Essays in the Theory of Risk-Bearing*, Amsterdam: North-Holland.
- ARUP. 2003. *Frameworks for Modelling the Variability of Journey Times on the Highway Network*. Arup, London.
- ARUP/OXERA. 2010. *What are the findings from the econometric analysis? Findings report*. Prepared for the Department for Transport, Transport Scotland and PDFC.
- ASENSIO, J. & MATAS, A. 2008. Commuters' valuation of travel time variability. *Transportation Research Part E: Logistics and Transportation Review*, 44, 1074-1085.
- ASSOCIATION OF TRAIN OPERATING COMPANIES. 2013. *Passenger Demand Forecasting Handbook Version 5.1*. London.
- BAGCHI, M. & WHITE, P. 2003. Use of public transport smartcard data for understanding travel behaviour. *European Transport Conference, Strasbourg, France*. Association for European Transport.
- BAGCHI, M. & WHITE, P. R. 2005. The potential of public transport smart card data. *Transport Policy*, 12, 464-474.
- BATES, J. 2009. An Agenda for Research on Reliability. *European Transport Conference, Leiden, Netherlands*. Association for European Transport.
- BATES, J., POLAK, J., JONES, P. & COOK, A. 2001. The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, 37, 191-229.
- BATLEY, R. 2007. Marginal valuations of travel time and scheduling, and the reliability premium. *Transportation Research Part E: Logistics and Transportation Review*, 43, 387-408.

- BATLEY, R.P., DARGAY, J.M., WARDMAN, M. 2011. The impact of lateness and reliability on passenger rail demand. *Transportation Research Part E: Logistics and Transportation Review*, 47, 61–72
- BATLEY, R., GRANT-MULLER, S., NELLTHORP, J., DE JONG, G., WATLING, D., BATES, J., HESS, S. & POLAK, J. 2008. *Multimodal Travel Time Variability*. Institute for Transport Studies. Leeds, UK
- BATLEY, R. & IBÁÑEZ, J. N. 2012. Randomness in preference orderings, outcomes and attribute tastes: An application to journey time risk. *Journal of Choice Modelling*, 5, 157-175.
- BATLEY, R.P., WARDMAN, M., IBANEZ, N., SHIRES, J. & WHELAN, G. 2006. *Revealed Preference Study to Assess Impact of Reliability on Passenger Demand. Deliverable 3: Report of Discrete Choice Analysis*. Prepared for Department of Transport.
- BEAUD, M., T. BLAYAC & M. STEPHAN, 2016. The impact of travel time variability and travellers risk attitudes on the values of time and reliability. *Transportation Research Part B: Methodological*, 93, 207-224.
- BENWELL M. & BLACK, I., 1984, *Train Service Reliability on BR InterCity Services*. Report 3: Passenger Attitudes to Lateness. London: Cranfield University.
- BERNOULLI, D. 1954. Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22, 23-36.
- BESSELING, P., DE GROOT, W. & VERRIPS, A. 2004. *Economische toets op de Nota Mobiliteit*, CPB document 65, CPB, The Hague.
- BHAT, C. R. & SARDESAI, R. 2006. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological*, 40, 709-730.
- BIERLAIRE, M. 2003. BIOGEME: A free package for the estimation of discrete choice models , *Proceedings of the 3rd Swiss Transportation Research Conference*. Ascona, Switzerland.
- BIERLAIRE, M. 2006. A theoretical analysis of the cross-nested logit model, *Annals of Operations Research*, 144, 287-300.
- BLACK, I. G., TOWRISS, J. G. 1993. Demand effects of travel time reliability. Final report prepared for London Assessment Division, UK Department for Transport.
- BLYTHE, P. 1998. Integrated Ticketing-Smart Cards in Transport. *IEE Colloquium: Using ITS in Public Transport and Emergency Services, Paper No. 4*

- BLYTHE, P. T. 2004. Improving public transport ticketing through smart cards. *Municipal Engineer*, 157, 47–54.
- BOGERS, E., VAN LINT, H. & VAN ZUYLEN, H. 2008. Reliability of Travel Time: Effective Measures from a Behavioral Point of View. *Transportation Research Record: Journal of the Transportation Research Board*, 2082, 27-34.
- BOGERS, E., VITI, F., HOOGENDOORN, S. & VAN ZUYLEN, H. 2007. Valuation of Different Types of Travel Time Reliability in Route Choice: Large-Scale Laboratory Experiment *Transportation Research Record: Journal of the Transportation Research Board*, 1985, 162-170.
- BÖRJESSON, M., 2008. Joint RP-SP data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E: Logistics and Transportation Review*, 44 (6), 1025–1038.
- BÖRJESSON, M., ELIASSON, J. & FRANKLIN, J. P. 2012. Valuations of travel time variability in scheduling versus mean–variance models. *Transportation Research Part B: Methodological*, 46, 855-873.
- BRADLEY, M. & DALY, A. 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation*, 21(2), 167-184.
- BROWNSTONE, D. & SMALL, K. A. 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39, 279-293.
- CARRION, C. & LEVINSON, D. 2010. Value of reliability: high occupancy toll lanes, general purpose lanes, and arterials. *Conference Proceedings of 4th International Symposium on Transportation Network Reliability*. Minneapolis, MN (USA).
- CARRION, C. & LEVINSON, D. 2012. Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, 46, 720-741.
- CARRION, C. & LEVINSON, D. 2013. Valuation of Travel Time Reliability from a GPS-based Experimental Design. *Transportation Research Part C: Emerging Technologies*, 35. 305–323.
- CASELLO, J., NOUR, A. & HELLINGA, B. 2010. Quantifying Impacts of Transit Reliability on User Costs *Transportation Research Record: Journal of the Transportation Research Board*, 2112/2009, 136-141.
- CHAKIROV, A. & ERATH, A. 2011. Use of public transport smart card fare payment data for travel behaviour analysis in Singapore. *16th International Conference of Hong Kong Society for Transportation Studies*. Hong Kong.
- CHAM, L. 2006. *Understanding Bus Service Reliability: A Practical Framework Using AVL/APC Data*. Master's Thesis. Massachusetts Institute of Technology. USA.

- CHAN, J. 2007. *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data*. Master's Thesis. Massachusetts Institute of Technology. USA.
- CHEN, X., YU, L., ZHANG, Y. & GUO, J. 2009. Analyzing urban bus service reliability at the stop, route, and network levels. *Transport Research Part A: Policy and Practice*, 43, 722-734.
- CHOI, J., LEE, Y., KIM, T. & SOHN, K. 2012. An analysis of Metro ridership at the station-to-station level in Seoul. *Transportation*, 39, 705-722.
- CHRISTOS, T. 2011. *Evaluation of the performance of London's Underground network with the use of smart card data records*. MSc Dissertation, University of Leeds. UK.
- COPLEY, G. MURPHY, P., PEARCE, D., 2002. Understanding and valuing journey time variability. *Proceedings of the European Transport Conference, Cambridge*. PTRC, London.
- DEATON, A. & MUELLBAUER, J., 1980. *Economics and Consumer Behaviour*. Cambridge University Press. Cambridge.
- DE JONG, G. & BLIEMER, M. C. J. 2015. On including travel time reliability of road traffic in appraisal. *Transportation Research Part A: Policy and Practice*, 73, 80-95.
- DE JONG, G., KOUWENHOVEN, M., KROES, E., RIETVELD, P. & WARFFEMIUS, P. 2009. Preliminary Monetary Values for the Reliability of Travel Times in Freight Transport. *European Journal of Transport and Infrastructure Research*, 9, 83-99.
- DE JONG, G., KOUWENHOVEN, M., BATES, J., KOSTER, P., VERHOEF, E., TAVASSZY, L. and WARFFEMIUS, P., 2014. New SP-values of time and reliability for freight transport in the Netherlands. *Transportation Research Part E: Logistics and Transportation Review*, 64, 71-87.
- DEPARTMENT FOR TRANSPORT, 2014. *WebTAG Unit A1.3: User and Provider Impacts*. London, UK.
- DEPARTMENT FOR TRANSPORT, 2015. *Values of travel time savings and reliability: final reports*. London, UK.
- DEPARTMENT FOR TRANSPORT, 2017. *WebTAG Unit A1.3: User and Provider Impacts*. London, UK.
- DEPARTMENT OF THE ENVIRONMENT, TRANSPORT AND THE REGIONS. 1998. *A New Deal for Transport—Better for Everybody*. London, UK.
- EDDINGTON, R. 2006. *The Eddington Transport Study*. HM Treasury. London, UK

- EHREKE, I., HESS, S., WEIS, C. & AXHAUSEN, K W. 2015. Reliability in the German value of time study. *Transportation Research Record*, 2495. 14-22.
- EL-GENEIDY, A. M., HORNING, J. & KRIZEK, K. J. 2010. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation*. 45. 66-79.
- ETTEMA, D. & TIMMERMANS, H. 2006. Costs of travel time uncertainty and benefits of travel time information: Conceptual model and numerical examples. *Transportation Research Part C: Emerging Technologies*, 14, 335-350.
- FABER MAUNSELL & MOTT MACDONALD. 2007. *Rail overcrowding, reliability and frequency demand impacts project*. Final report to Centro.
- FISHBURN, P. 1970. *Utility Theory for Decision Making*. Wiley, New York.
- FOSGERAU, M. & ENGELSON, L. 2011. The value of travel time variance. *Transportation Research Part B: Methodological*, 45, 1-8.
- FOSGERAU, M. & FUKUDA, D. 2010. *Valuing travel time variability: Characteristics of the travel time distribution on an urban road*. MPRA Paper. University Library of Munich, Germany,
- FOSGERAU, M. & KARLSTRÖM, A. 2010. The value of reliability. *Transportation Research Part B: Methodological*, 44, 38-49.
- FRIEDMAN, M. & SAVAGE, L. J. 1948. The Utility Analysis of Choices Involving Risk. *Journal of Political Economy*, 56, 279-304.
- FRUMIN, M. 2008. *Oyster-Based Performance Metrics for the London Overground*. Master's Thesis Massachusetts Institute of Technology. USA.
- GAVER, D. P. 1968. Headstart Strategies for Combating Congestion. *Transportation Science*, 2, 172-181.
- GEISTEFELDT, J., HOHMANN, S. & WU, N. 2014. *Ermittlung des Zusammenhangs von Infrastruktur und Zuverlässigkeit des Verkehrsablaufs für den Verkehrsträger Straße*. Schlussbericht für Bundesministerium für Verkehr und digitale Infrastruktur. Ruhr Universität Bochum.
- GHOSH, A., 2001. Valuing time and reliability: commuters\_ mode choice from a real time congestion pricing experiment. Ph.D. Dissertation, Department of Economics, University of California at Irvine.
- GREATER LONDON AUTHORITY. 2010. *Mayor's Transport Strategy*. Available: <https://www.london.gov.uk/priorities/transport/publications/mayors-transport-strategy>
- HABER, A. & RUNYON, R. P. 1973. *General Statistics*, Reading, Mass.: Addison-Wesley.

HELLINGA, B. 2011. *Defining, Measuring, and Modelling Transportation Network Reliability*. Final report, Delft University of Technology, the Netherlands.

HENSHER, D. A. 2001. The valuation of commuter travel time savings for car drivers in New Zealand: Evaluating alternative model specifications. *Transportation*. 28. 101-118

HENSHER, D. A. 2010. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological*, 44(6), 735-752.

HESS, S. HENSHER, D.A. & DALY, A. 2012. Not bored yet – Revisiting respondent fatigue in stated choice experiments, *Transportation Research Part A: Policy and Practice*, 46 (3), 626-644.

HICKS, J.R. 1935. A Suggestion for Simplifying the Theory of Money. *Economica* February :1-19.

HOLLANDER, Y. 2006. Direct versus indirect models for the effects of unreliability. *Transportation Research Part A: Policy and Practice*, 40, 699-711.

JACKSON W.B. & JUCKER J.V. 1982. An empirical study of travel time variability and travel choice behavior. *Transportation Science*, Vol. 16 no4, 460 –475.

JENELIUS, E., MATTSSON, L.-G. & LEVINSON, D. 2011. Traveler delay costs and value of time with trip chains, flexible activity scheduling and information. *Transportation Research Part B: Methodological*, 45, 789-807.

JANG, W. 2010. Travel Time and Transfer Analysis Using Transit Smart Card Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2144, 142-149.

KAHNEMAN, D. & TVERSKY, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47, 263-291.

KEYNES, J. M., 1936. *The general theory of employment, interest and money*. Macmillan

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

KOUWENHOVEN, M., SCHOEMAKERS, A., GROU, R. VAN, KROES, E.P. 2005 Development of a tool to assess the reliability of Dutch road networks. *European Transport Conference 2005*, Strasbourg.

KOUWENHOVEN, M., DE JONG, G.C., KOSTER, P., VAN DEN BERG, V.A.C., VERHOEF, E.T., BATES, J., WARFFEMIUS, P.M.J. 2014. New values of time and reliability in passenger transport in the Netherlands, *Research in Transport Economics* 47, p.37-49.

- KNIGHT, F. H. 1921. *Risk, Uncertainty and Profit*, New York: Houghton Mifflin.
- KNIGHT, T. 1974. An approach to the evaluation of changes in travel unreliability: A “Safety margin” hypothesis. *Transportation*, 3, 393-408.
- KURAUCHI, F., SCHMÖCKER, J.D., SHIMAMOTO, H., & HASSAN, S. M. 2014. Variability of commuters’ bus line choice: An analysis of oyster card data. *Public Transport*, 6, 21-34.
- KUSAKABE, T. & ASAKURA, Y. 2014. Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies*, 46, 179–191.
- LAM, T. C. & SMALL, K. A. 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review*, 37, 231-251.
- LEE, S.G., & HICKMAN, M. 2014. Trip purpose inference using automated fare collection data. *Public Transport*. 6. 1-20.
- LEVINSON, D. & ZHU, S. 2013. A portfolio theory of route choice. *Transportation Research Part C: Emerging Technologies*, 35, 232-243.
- LEVY, H. & MARKOWITZ, H. M. 1979. Approximating Expected Utility by a Function of Mean and Variance. *The American Economic Review*, 69, 308-317.
- LI, H., TU, H. and HENSHER, D.A., 2016. Integrating the mean–variance and scheduling approaches to allow for schedule delay and trip time variability under uncertainty. *Transportation Research Part A: Policy and Practice*, 89, 151-163.
- LI, Z., HENSHER, D. A. & ROSE, J. M. 2010. Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation Review*, 46, 384-403.
- LIU, H. X., HE, X., RECKER, W., 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research Part B: Methodological*, 41, 448–461.
- LIU, H. X., RECKER, W. & CHEN, A. 2004. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice*, 38, 435-453.
- LU, H., FOWKES, A.S., WARDMAN, M. 2008. Amending the incentive for strategic bias in stated preference studies: a case study in the users’ valuation of rolling stock. *Transportation Research Record*. 2049, 128-135.



- MA, X., WU, Y.J., WANG, Y., CHEN, F. & LIU, J. 2013. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36, 1-12
- MAHMASSANI, H. & STEPHAN, D. 1988. Experimental Investigation of Route and Departure Time Choice Dynamics of Urban Commuters. *Transportation Research Record*. 1203, 69-84.
- MAHMASSANI, H.S. 2011. *Incorporating reliability performance measures in planning and operations modelling tools*, SHRP2 L04 report, Annual Meeting of the Transportation Research Board, Washington DC.
- MARKOWITZ, H. M. 1952, Portfolio Selection. *The Journal of Finance*, 7: 77–91.
- MARKOWITZ, H. M. 1959. *Portfolio selection: efficient diversification of investments*, London, Chapman & Hall.
- MARSCHAK, J. 1938. Money and the Theory of Assets. *Econometrica*, vol. 6:311-325.
- MAZLOUMI, E., CURRIE, G. & ROSE, G. 2008. Causes of travel time unreliability – a Melbourne case study. *Australasian Transport Research Forum 2008 Proceedings*.
- MAZLOUMI, E., CURRIE, G. & ROSE, G. 2010. Using GPS Data to Gain Insight into Public Transport Travel Time Variability. *Journal of Transportation Engineering*. 136. 623-631.
- MCFADDEN, D. 1974. Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (ed.) *Frontiers in Econometrics*. Academic Press, New York.
- MINISTRY OF TRANSPORT, WATERWORKS AND INFRASTRUCTURE. 2004, *Nota Mobiliteit.Naar een betrouwbare en voorspelbare bereikbaarheid, deel 1*, The Hague, The Netherlands.
- MORENCY, C., TREPANIER, M. & AGARD, B. 2006. Analysing the Variability of Transit Users Behaviour with Smart Card Data. *Intelligent Transportation Systems Conference*, 17-20 Sept. 2006 44-49.
- MVA. 1989. *Network SouthEast quality of service phase III final report*. Prepared for Network SouthEast, British Railways Board.
- MVA. 2000. *Valuation of reliability and punctuality improvements on rail services*. Final Report to Shadow Strategic Rail Authority.
- MVA. 2008. *Econometric analysis of long time series rail passenger demand aggregates*. Prepared for Department for Transport, UK.
- MVA. 2009. *Regional flows: regional rail demand elasticities*. Report to Passenger Demand Forecasting Council.

MVA & ITS LEEDS. 2012. *Demand impacts of recovery time*. Report to Passenger Demand Forecasting Council, ATOC, London.

NASSIR, N., HICKMAN, M. & MA, Z.-L. 2015. Behavioural findings of observed transit route choice strategies from the farecard data in Brisbane. *Australasian Transport Research Forum 2015 Proceedings*.

NOLAND, R. B. & POLAK, J. W. 2002. Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, 22, 39-54.

NOLAND, R. B. & SMALL, K. A. 1995. Travel-Time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes. *Transportation Research Record: Journal of the Transportation Research Board*, 150-158.

ORR, S. HESS, S & SHELDON, R. 2012. Consistency and fungibility of monetary valuations in transport: an empirical analysis of framing and mental accounting effects. *Transportation Research Part A: Policy and Practice*, 46 (10). 1507 - 1516.

ORTÚZAR, J.D. 2001. On the development of the nested logit model, *Transportation Research* 35B, 213-216.

ORTÚZAR, J.D. & WILLUMSEN, L.G. 2011. *Modelling Transport*, 4th edition 466-468. West Sussex, U.K.: John Wiley and Sons Ltd.

PARK, J., KIM, D.-J. & LIM, Y. 2008. Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea. *Transportation Research Record: Journal of the Transportation Research Board*, 2063, 3-9.

PARLIAMENTARY OFFICE OF SCIENCE AND TECHNOLOGY. 2014. Big and Open Data in Transport. POSTnotes POST-PN-472.

PEER, S., KOOPMANS, C. & VERHOEF, E.T. 2012. Prediction of travel time variability for cost-benefit analysis. *Transportation Research Part A: Policy and Practice*, 46, 79-90.

PELLETIER, M.-P., TRÉPANIÉ, M. & MORENCY, C. 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19, 557-568.

PELLS, S. 1987. The evaluation of reductions in travel time variability. PhD thesis in Economics, University of Leeds, UK.

POLAK, J., 1987a. *Travel time variability and departure time choice: a utility theoretic approach*. Discussion Paper No. 15, Transport Studies Group, Polytechnic of Central London, Great Britain.

POLAK, J., 1987b. A more general model of individual departure time choice. In: *PTRC Summer Annual Meeting*, Proceedings of Seminar C.

- PRATO, C., RASMUSSEN, T. & NIELSEN, O. 2014. Estimating value of congestion and of reliability from observation of route choice behavior of car drivers. *Transportation Research Record*.
- PRATT, J. W. 1964. Risk Aversion in the Small and in the Large. *Econometrica*, 32, 122-136.
- RICHTER, M. K. 1960. Cardinal Utility, Portfolio Selection and Taxation. *The Review of Economic Studies*, 27, 152-166.
- RIETVELD, P., BRUINSMA, F. R. & VAN VUUREN, D. J. 2001. Coping with unreliability in public transport chains: A case study for Netherlands. *Transportation Research Part A: Policy and Practice*, 35, 539-559.
- ROM, B. M. & FERGUSON, K. W. 1994. Post-modern portfolio theory comes of age. *The Journal of Investing*, 3(3), 11-17.
- ROSE, J.M. & HENSHER, D.A. 2014. Tollroads are only part of the overall trip: the error of our ways in past willingness to pay studies. *Transportation*, 41:4, 819-837
- SAMUELSON, P. A. 1970. The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Moments. *The Review of Economic Studies*, 37, 4, 537-542.
- SAVAGE, L. J. 1954. *The Foundations of statistics*, New York: Wiley.
- SCHIL, M. 2012. *Measuring Journey Time Reliability in London Using Automated Data Collection Systems*. Master's Thesis. Massachusetts Institute of Technology. USA.
- SEABORN, C., ATTANUCCI, J. & WILSON, N. 2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2121, 55-62.
- SENNA, L. D. S. 1994. The influence of travel time variability on the value of time. *Transportation*, 21, 203-228.
- SMALL, K. A. 1982. The Scheduling of Consumer Activities: Work Trips. *The American Economic Review*, 72, 467-479.
- SMALL K.A., NOLAND R., CHU X. & LEWIS D. 1999. *Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation*. National Cooperative Highway Research Program Report 431, National Academy Press
- SMALL, K., NOLAND, R., & KOSKENOJA, P. 1995. *Socio-economic attributes and impacts of travel reliability: a stated preference approach*. California Partners for Advanced Transit and Highways (PATH).
- SMALL, K. A., WINSTON, C. & YAN, J. 2005. Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica*, 73, 1367-1382.

- SORRATINI, J., LIU, R., & SINHA, S. 2008. Assessing Bus Transport Reliability Using Micro-Simulation, *Transportation Planning and Technology*, 31, 303-324.
- STEER DAVIES GLEAVE. 2003. *The effect of road congestion on rail demand*. Report to Passenger Demand Forecasting Council.
- STERMAN, B.P. & SCHOFER, J.L. 1976. Factors Affecting Reliability of Urban Bus Services. *Transportation Engineering Journal of ASCE*, 102, 147-159.
- STRATHMAN, J. & HOPPER, J. 1993. Empirical Analysis of Bus Transit On-Time Performance. *Transportation Research Part A: Policy and Practice*. 27. 93-100.
- TAVASSOLI, A., ALSGER, A., HICKMAN, M. & MESBAH, M. 2016. How close the models are to the reality? Comparison of Transit Origin-Destination Estimates with Automatic Fare Collection Data. *Australasian Transport Research Forum 2016 Proceedings*.
- TILAHUN, N & LEVINSON, D. 2010. A moment of time: reliability in route choice using stated preference. *Journal of Intelligent Transportation Systems*, 14, 179–187
- TRANSPORT FOR LONDON. 2011. *Developing a reliability metric for LU customers*. Unpublished. TfL, London.
- TOBIN, J. 1958. Liquidity Preference as Behavior towards Risk. *Review of Economic Studies*, 25, 1:65-86.
- TPA. 1992. *Trans-Pennine rail strategy study: model calibration and validation report*. Prepared for GMPTE, Merseytravel, SYPT, WYPTE, Cheshire CC, Derbyshire CC, Humberside CC, Lancashire CC and Peak Park.
- TRAIN, K. E. 2009. *Discrete choice methods with simulation*. 2<sup>nd</sup> Edition, Cambridge University Press.
- TSENG, Y.-Y., VERHOEF, E., DE JONG, G., KOUWENHOVEN, M. & VAN DER HOORN, T. 2009. A pilot study into the perception of unreliability of travel times using in-depth interviews. *Journal of Choice Modelling*, 2, 8-28.
- TSIANG, S. C. 1972. The Rationale of the Mean-Standard Deviation Analysis, Skewness Preference, and the Demand for Money. *The American Economic Review*, 62, 354-371.
- UNIMAN, D. 2009. *Service reliability measurement framework using smart card data : application to the London Underground*. Master's Thesis. Massachusetts Institute of Technology. USA.
- UTSONOMIYA, M., ATTANUCCI, J. & WILSON, N., 2006. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1971, 119–126.

VAN LINT, J. W. C., VAN ZUYLEN, H. J. & TU, H. 2008. Travel time unreliability on freeways: Why measures based on variance tell only half the story. *Transportation Research Part A: Policy and Practice*, 42, 258-277.

VICKREY, W. S. 1969. Congestion Theory and Transport Investment. *The American Economic Review*, 59, 251-260.

VON NEUMANN, J. & MORGENSTERN, O. 1947, *Theory of games and economic behavior*. Princeton, NJ: Princeton University.

WARDMAN, M. 1988. A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour. *Journal of Transport Economics and Policy*, 22, 71-91.

WARDMAN, M. & BATLEY, R. 2014. Travel time reliability: a review of late time valuations, elasticities and demand impacts in the passenger rail market in Great Britain. *Transportation*, 1-29.

WEI, W. 2010. *Bus passenger origin-destination estimation and travel behavior using automated data collection systems in London, UK*. Master's Thesis. Massachusetts Institute of Technology. USA.

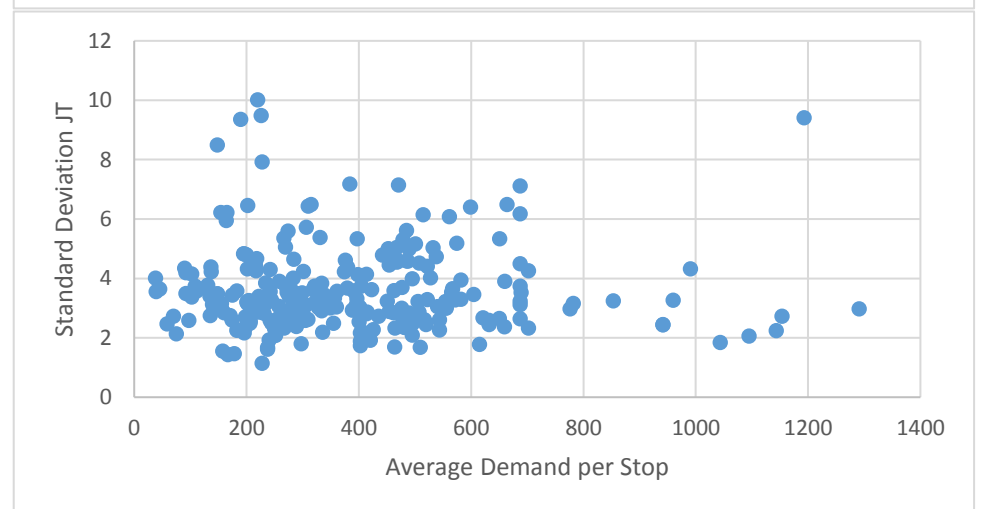
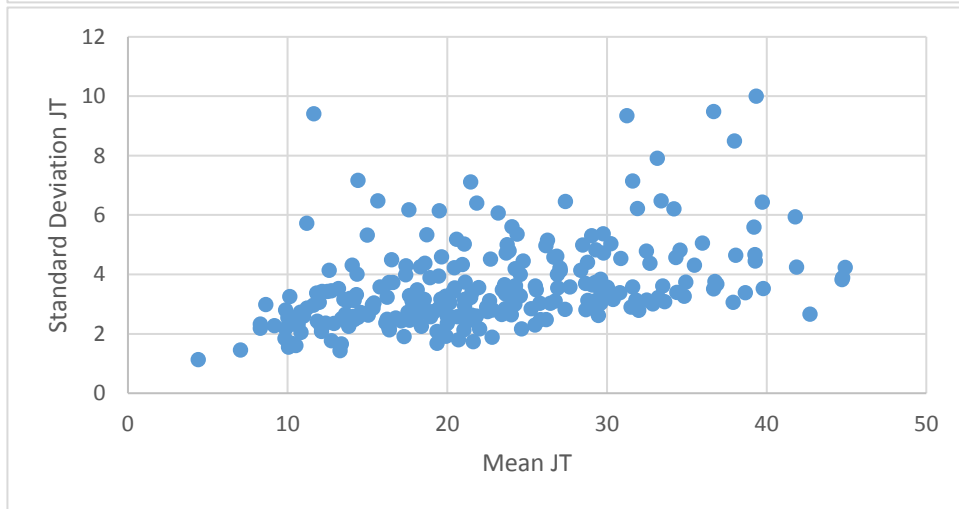
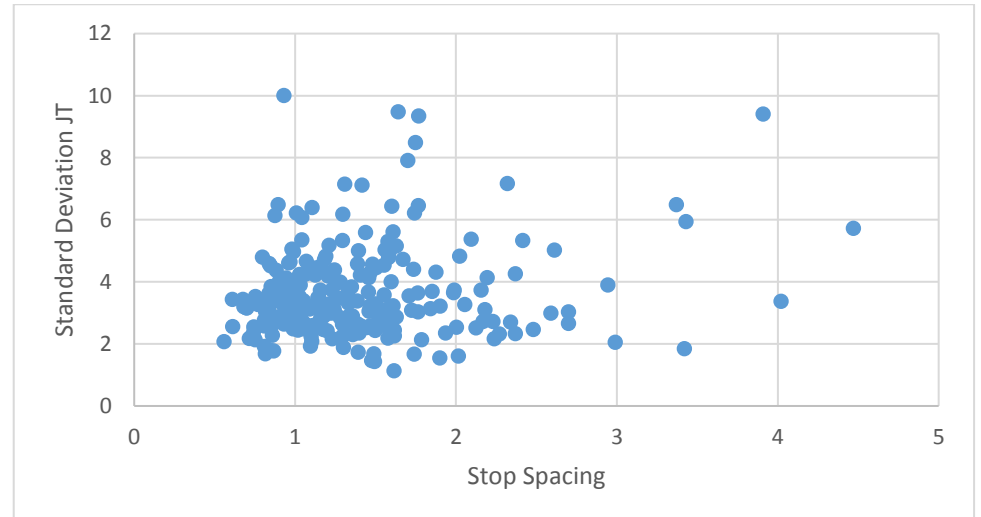
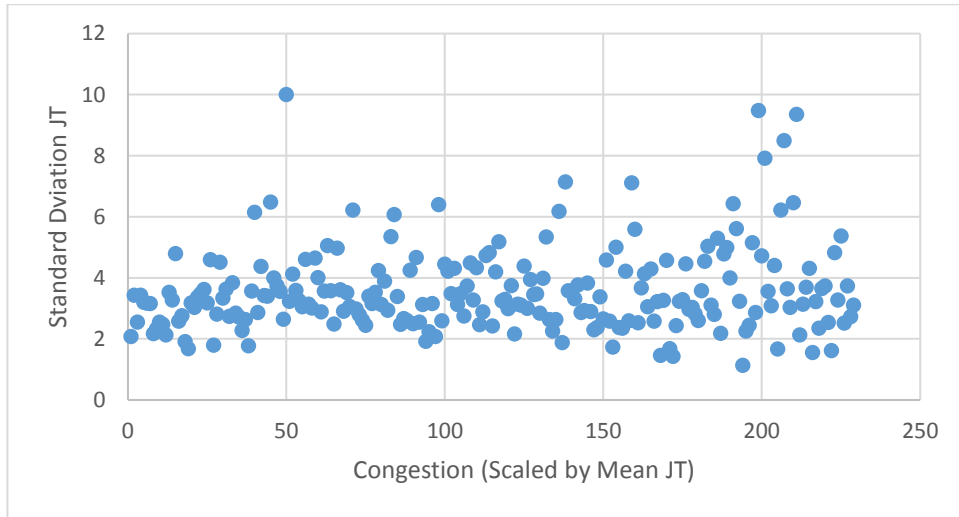
WEN, C.-H. & KOPPELMAN, F. S. 2001. The generalized nested logit model. *Transportation Research Part B: Methodological*, 35, 627-641.

WHITE, P., BAGCHI, M. & BATAILLE, H. 2010. The role of smart card data in Public Transport. *12th WCTR*. Lisbon, Portugal.

XU, H. 2007. Evaluation of benefits and effectiveness of smart cards for public transport. PhD thesis in Transport Studies, University of Leeds, UK.



Appendix 1 - Scatter plots of statistically significant explanatory variables against standard deviation of journey time from Table 4.5



Appendix 2 – Variables used to estimate discrete choice models in Chapter 5

Origin Destination	Mode 1								Mode 2					
	Mode 1	Mode 2	Count	Mean	Median	St Dev	Transfer	Headway	Count	Mean	Median	St Dev	Transfer	Headway
Arnos Grove Canary Wharf	LU	LU/DLR	62.00	55.48	57.00	13.19	1.00	4.43	20.00	55.55	55.00	7.04	1.00	14.43
Woodside Park Canary Wharf	LU	LU/DLR	52.00	55.44	55.00	5.07	1.00	7.86	40.00	56.00	55.50	3.65	1.00	6.00
Morden Canary Wharf	LU	LU/DLR	142.00	50.35	49.00	6.04	1.00	4.71	22.00	52.91	52.00	4.16	1.00	6.71
East Finchley Canary Wharf	LU	LU/DLR	27.00	47.22	47.00	6.31	1.00	8.43	24.00	49.25	49.00	3.64	1.00	10.43
Debden Canary Wharf	LU	LU/DLR	54.00	46.35	45.00	5.19	1.00	7.86	26.00	46.77	47.50	3.51	1.00	6.00
Colliers Wood Canary Wharf	LU	LU/DLR	69.00	44.94	43.00	5.20	1.00	4.71	19.00	49.21	48.00	4.30	1.00	6.71
Hornchurch Canary Wharf	LU	LU/DLR	85.00	44.41	43.00	5.57	1.00	6.00	23.00	43.00	42.00	7.05	1.00	8.00
Hampstead Canary Wharf	LU	LU/DLR	30.00	44.17	43.00	4.96	1.00	8.43	39.00	44.92	45.00	3.25	1.00	10.43
Kentish Town Canary Wharf	LU	LU/DLR	19.00	38.95	39.00	3.76	1.00	8.43	32.00	37.38	36.50	4.21	1.00	10.43
Barkingside Canary Wharf	LU	LU/DLR	37.00	38.70	38.00	3.84	1.00	4.29	15.00	41.67	42.00	1.95	1.00	6.00
Buckhurst Hill Canary Wharf	LU	LU/DLR	119.00	38.50	37.00	6.85	1.00	4.29	15.00	41.67	41.00	6.24	1.00	9.57
Shepherds Bush Canary Wharf	LU	LU/DLR	96.00	37.47	36.00	5.56	1.00	4.00	28.00	43.25	42.00	3.30	1.00	6.00
Clapham South Canary Wharf	LU	LU/DLR	355.00	36.50	35.00	6.06	1.00	4.71	43.00	41.65	40.00	5.79	1.00	6.71
Gants Hill Canary Wharf	LU	LU/DLR	143.00	33.21	32.00	5.44	1.00	7.86	34.00	37.21	37.00	4.26	1.00	6.00
South Woodford Canary Wharf	LU	LU/DLR	226.00	31.78	31.00	5.16	1.00	7.86	73.00	35.08	34.00	4.37	1.00	6.00
Snaresbrook Canary Wharf	LU	LU/DLR	100.00	31.50	30.00	6.21	1.00	7.86	23.00	34.35	33.00	5.84	1.00	6.00
Old Street Canary Wharf	LU	LU/DLR	71.00	23.03	22.00	4.50	1.00	8.43	50.00	24.58	24.00	3.60	1.00	10.43
Leyton Canary Wharf	LU	LU/DLR	179.00	22.08	21.00	5.06	1.00	7.86	34.00	26.21	26.00	4.05	1.00	6.00
Ealing Broadway Kings Cross	LU	LU/Rail	49.00	41.67	41.00	8.17	0.00	8.00	35.00	42.86	42.00	5.36	1.00	15.00
Paddington Old Street	LU	LU/Rail	25.00	30.68	33.00	13.93	1.00	7.71	22.00	35.45	34.00	5.10	1.00	20.00
Walthamstow Central Farringdon	LU	LU/Rail	108.00	32.79	32.00	4.53	1.00	2.00	15.00	35.40	36.00	4.42	1.00	12.00
Tooting Broadway Victoria	LU	LU/Rail	280.00	27.00	26.00	5.35	1.00	4.71	16.00	28.75	28.50	2.91	1.00	8.71
Vauxhall Canary Wharf	LU	LU/Rail	37.00	31.14	30.00	5.20	1.00	4.71	73.00	28.67	27.00	5.89	1.00	10.00
Seven Sisters Bethnal Green	LU	LU/Rail	20.00	37.15	36.50	2.30	1.00	4.00	23.00	26.74	26.00	5.51	1.00	17.00
Harrow On The Hill Marvlebone	LU	LU/Rail	19.00	33.89	33.00	4.54	1.00	8.29	34.00	21.21	19.50	5.04	1.00	30.00
Brixton Victoria	LU	Rail	546.00	13.73	13.00	2.53	0.00	2.00	23.00	12.13	10.00	4.73	0.00	15.00
Richmond Victoria	LU	LU/Rail	46.00	39.15	39.00	4.59	0.00	9.00	68.00	28.91	28.00	6.28	1.00	21.00
Motspur Park Victoria	Rail	LU/Rail	19.00	33.37	32.00	5.97	1.00	25.00	20.00	38.20	37.50	3.35	1.00	17.00
Wimbledon London Bridge	Rail	LU/Rail	16.00	38.63	36.00	8.21	1.00	12.00	19.00	36.47	34.00	8.83	1.00	8.00
Earlsfield London Bridge	Rail	LU/Rail	18.00	34.61	35.00	3.22	1.00	11.00	22.00	31.23	30.00	5.75	1.00	6.00
Clapham Junction London Bridge	Rail	LU/Rail	50.00	29.50	29.50	4.01	1.00	15.00	111.00	26.07	24.00	5.34	1.00	10.00



Appendix 3 – XY plots of relationships between explanatory variables

