The economics of rail passenger responses to delays: delay perception, satisfaction, valuation and compensation

Kacper Rossa

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

The University of Leeds

Institute for Transport Studies

June 2025

List of publications and conferences attended for this PhD

The following journal articles and technical reports are based on the analysis presented in the thesis:

Journal papers

Rossa et al. (2024.). The problem of homogeneity of rail passenger delay compensation scheme rules in Great Britain: impacts on passenger engagement and operator revenues. *European Transport Research Review*. **16**(15). This work is based on Chapter 4.

Rossa et al. (2024). The valuation of delays in passenger rail using journey satisfaction data. *Transportation Research Part D: Transport and Environment* **129**. 104088. This work is based on Chapter 8.

Technical report

Batley, R., Rossa, K. Smith, A. (2023). Passengers' acceptability of delays. Technical report and advice to Network Rail regarding passengers' acceptability of delays. This work was based on Chapter 7 (co-authored with Richard Batley and Andrew Smith).

Work from this thesis was presented at following conferences:

Transport Research Arena 2022. *The problem of homogeneity of rail passenger delay compensation scheme rules in Great Britain: impacts on passenger engagement and operator revenues*. Lisbon, Portugal. November, 2022.

Young Researchers Seminar 2023 (YRS 2023). *Rail delays and travellers' perception of being delayed*. Lisbon, Portugal. May 2023.

International Transportation Economics Association (ITEA) Annual Conference 2023. P*assengers' perception of delays, its impacts on satisfaction and lateness valuation using data on passenger satisfaction*. Santander, Spain. June 2023.

11th Symposium of the European Association for Research in Transportation (hEART 2023). *The valuation of arrival and departure delays in the UK passenger rail using satisfaction survey data*. Zürich, Switzerland. September 2023.

Acknowledgements

I would like to acknowledge the Engineering and Physical Sciences Research Council and Arup for funding this PhD. I would like to thank my supervisors at the Institute for Transport Studies at the University of Leeds, Andrew Smith and Richard Batley for all their help, encouragement and support throughout the way. Further thanks to Phil Hudson from Arup for his industrial insights and help with organising the work placement at Arup. Here, I would also like to extend my gratitude to the whole Arup team and especially Margot Finley, Josie Drath, Iris Ning, Kieran Arter, Ryan Liley, Chris Prior and Pawel Bugajski for making me feel welcome at Arup.

I would also like to express my gratitude to Transport Focus and especially David Greeno, Murray Leader and Toby Cotton for sharing the NRPS data and all the support they provided.

My special thanks also go to Adriaan Roeleveld and Mark Wardman for finding time to discuss rail delays.

I would also like to thank all the colleagues I met at all the conferences I attended througouht this PhD. Thank you for listening to my presentations and all the valuable suggestions. Here, I would especially like to thank Ingrid van Schagen for her help and mentoring in preparing submission to YRS.

Many thanks for all the discussions with fellow PhD students and academic staff at ITS at various stages of my education, especially the Economics and Appraisal Group as well as MSc Mathematical Modelling for Transport with special thanks to David Watling and Richard Connors who supported me in the application process as well my MSc thesis supervisor Tom van Vuren.

I also extend my thanks to my previous team at Mott MacDonald, especially Jon Crockett and Andrew Gillies-Smith for all I learned while working with them.

Lastly, I would also like to thank my family and friends, and especially my partner, Sabina for the continuous support.

Abstract

Though delays negatively impact rail passengers, eliminating all their incidences is impossible. Hence, it is important to understand the impacts of delays of various lengths and ways in which passengers can be compensated for the resulting disutility. This can, in turn, help dictate regulatory, operational and investment priorities. Against this background, this thesis aims to review the currently operating delay compensation mechanisms and investigate the link between delay occurrence, delay perception and satisfaction impacts. The results suggest that the current compensation scheme rules lead to an increasing revenue burden of the scheme for long-distance operators, highlighting the need for further research comparing the scheme's costs and benefits. Subsequently, it was indicated that shorter delays are not always perceived by passengers and are likely to have a smaller impact on passenger satisfaction with marginal disutility of delay likely being non-constant across the different delay levels. At the same time, it was highlighted that journey quality, delay at departure and journey time also affect delay perception and satisfaction. The probability of perceiving a delay was estimated to be larger than the equivalent probability of being dissatisfied with the same delay, demonstrating the existence of a gap between delay perception and dissatisfaction. Finally, the journey satisfaction data were used to derive lateness multipliers, a conversion rate between a minute of delay and scheduled journey time. The calculated values were found to be larger than the estimates obtained from the traditionally used stated preference studies. The outcomes of the research conducted as part of this thesis can help design passenger delay compensation schemes and devise performance strategies and targets for railways. Moreover, the presented analysis provides additional evidence towards possible non-linearities in delay impacts and highlights the potential of transport satisfaction data in economic valuation.

Contents

List of Figures

List of Tables

List of abbreviations

AML	Average Minutes Late
APF	Die Agentur für Passagier- und Fahrgastrechte
APL	Average Passenger Lateness
AVE	Alta Velocidad Española
COVID	Coronavirus Disease
CRR	Compensation to Revenue Ratio
DfT	Department for Transport
DR	Delay Repay
ERA	European Regions Airline Association
EU	European Union
FE	Fixed Effects
GB	Great Britain
GJT	Generalised Journey Time
HSP	Historic Service Performance
IM	Infrastructure Manager
LCC	Low Cost Carrier
LNER	London North Eastern Railway
LSE	London South East
NEXTOR	National Center of Excellence in Aviation Operations Research
NRPS	National Rail Passenger Survey
ÖBB	Österreichische Bundesbahnen
OD	Origin-Destination
OLS	Ordinary Least Squares
ORR	Office of Rail and Road
PAT	Preferred Arrival Time
PDFH	Passenger Demand Forecasting Handbook
PPM	Public Performance Measure
PSO	Public Service Obligation
RE	Random Effects
RID	Rail Identifier
RP	Revealed Preference
SJT	Scheduled Journey Time
SP	Stated Preference
TfW	Transport for Wales

TOC	Train Operating Company
TRUST	Trains Running Under System TOPS (Total Operation Processing System)
UK	United Kingdom
UTK	Urząd Transportu Kolejowego

Chapter 1
Introduction

1.1. Background

In the fast-paced world we live in, it has become of great importance to be able to move quickly. Transport is an important part of everyday life, being a means of moving between different places, required to access jobs, goods, services and leisure activities (Glazener et al., 2021). Different modes of transport have their specific characteristics, meaning that their suitability depends on the type of trip made and personal preferences. Similar to other areas of economics, the monetary cost of travel impacts the demand for it. However, in the case of transport, travellers simultaneously try to minimise the monetary cost and journey times. Whilst in some cases, travel activity on its own may be a source of enjoyment and/or fitness in the case of active modes (Cornet et al., 2022; Mokhtarian & Salomon, 2001; Wardman & Lyons, 2016), usually both increased travel time and cost are sources of disutility. Hence, transport is typically called a derived demand as it is related to the need to move between places, not to the travel activity itself.

With travel time being a source of negative utility, it is worth mentioning that in the case of public transport users, the time spent travelling is not only related to the time spent in-vehicle. In other words, public transport trips also include time spent accessing stops or stations as well as waiting for the chosen services. Finally, whilst both private and public transport users are concerned with the length of time their trips are meant to take (i.e. based on their expectations or timetables), congestion and resulting delays have been suggested to affect passengers more than changes in scheduled times. In other words, a minute of delay is considered to be worse than an additional minute of scheduled journey time. This may be related to the inconvenience and uncertainty caused by delays, the possibility of interfering with other planned activities and subsequently the financial consequences of missed connections, appointments etc. However, the amount of inconvenience (or disutility) derived from the additional travel time resulting from a delay may depend on how sensitive a traveller is to potential late-running and the experienced travel conditions.

1.1.1. Journey scheduling

When choosing a service to travel on, passengers consider the scheduled arrival time of the services with respect to their preferred arrival time. As noted by Preston et al. (2009) and explained in more detail in Batley (2007), passengers typically also include safety buffers to their travel schedules that work as a time insurance against any possible disruptions and aim to increase the probability of arriving to the destination within passenger’s preferred arrival time. The amount of buffer time may depend on the expectations, sensitivity to potential lateness, and personal characteristics of a passenger (this is further revisited in Chapter 2).

1.1.2. Impacts of delays on passengers

The impacts of delays on rail passengers are typically studied in terms of how performance affects demand. Intuitively, it would be expected that fewer trips will be made if the incidence of delays increases. Such a relationship is generally suggested by the literature, however, the estimated elasticities with respect to delay are inelastic. Performance has a statistically significant, yet marginal impact on rail demand as the estimated elasticities are typically not more (in absolute terms) than -0.10 (ATOC, 2004; Batley et al., 2011). At the same time, the relative impacts of changes in price levels or generalised journey times are suggested to be larger with the estimated elasticities closer to -1, indicating that demand is relatively responsive to such changes (ATOC, 2004; Batley et al., 2011). The limited impact of performance on demand does not necessarily mean that travellers place low value on performance as individual-level studies suggest relatively high valuation of lateness, yet such experiences do not always lead to a demand response (Batley et al., 2011). This may be related to passengers not being able to change their travel behaviour following late running and especially so in the short term, which can be attributed to the lack of available alternatives. This means that while delays may have a negative impact on passengers, performance may not always be immediately linked to demand. Nevertheless, delays do contribute to inconvenience and the loss of time that could have been spent on other activities. Hence, this could be translated to a loss of social welfare resulting from increased travel times.

1.1.3. Alternative methodologies used to study the impacts of delay

With studies of demand not being able to capture the aforementioned impacts, some alternative lines of investigation have been used throughout the literature, employing the following data sources:

1) stated preference (SP) surveys where respondents choose the preferred travel options based on the presented scenarios with different travel attributes (e.g. travel time and monetary cost),
2) revealed preference (RP) surveys where travellers' real choices are observed and compared to the alternatives, and
3) satisfaction surveys where passengers report journey satisfaction based on their travel experiences.

Demand and SP studies typically focus on evaluating the impacts of average performance (e.g. Batley and Ibáñez, 2012) on passenger numbers or preferences (this is further discussed in Chapter 8). At the same time, studies of passenger satisfaction are concerned with either overall satisfaction with public transport (e.g. Cats et al., 2015) or the impact of delay incidents (or other travel attributes) on passenger satisfaction with a specific travel experience (e.g. Monsuur et al., 2021) (Chapter 6 and Chapter 7 cover these aspects in more detail). Studying the impact of average performance on demand can provide information for the government and/or operators regarding the effects of changes in rail performance on passenger numbers and ticket revenues. Stated preference surveys are often used to estimate the relative importance of different travel attributes and trade-offs between them. These, are, however, related to hypothetical choices. On the other hand, studying real choices obtained through questionnaires and travel diaries also has its limitations as it assumes that travellers know about the other travel alternatives. Moreover, in the case of delays, these often cannot be predicted and are unlikely to affect any travel choices once a journey has started. Hence, the other body of literature is focused on studying the impact of different journey attributes on passenger satisfaction. In this case, passengers evaluate their actual experience *ex post*, giving insight into how different travel attributes (including delays) affected their satisfaction with the experienced journey.

Several studies looked at the impact of different aspects of the journey on travel satisfaction (for reviews see De Vos et al., 2013; De Oña and De Oña, 2015; Gao et al., 2018; Rong et al., 2022), but studies relating actual performance to satisfaction are more limited. In principle, this thesis draws on a number of previous studies examining the impacts of lateness on passengers (i.e. Batley, 2007; Preston et al., 2009; Monsuur et al., 2021) whilst also responding to the conclusion in Wardman and Batley (2022) and Rong et al. (2022) that further research is needed to understand passengers' perception of late-running and its marginal impacts on passengers. This is seen as an important first step in advancing our knowledge of the impact that delays have on passengers' satisfaction and the role that perception has in this relationship.

1.1.4. Focus on British railways

The main focus of this thesis is on the railways in Great Britain. This can be attributed to two reasons. First of all, the fact that the British rail planning practice is well-established with the first edition of British guidance on rail demand forecasting dating back to the 1980s (Wheat and Wardman, 2017). Secondly, the British railway industry structure is one characterised by vertical separation and privatisation of train operations (Nash et al., 2013), though in recent years a number of services has been brought back under the government's control and are run by the so-called 'operator of last resort'. With multiple train operators providing services across the country, extensive regulation and a relatively complex industry structure, the British rail industry is also characterised by a relative abundance of open-access data what facilitates research. Yet, there still remain areas that call for further research.

In November 2019, 65% of rail station stops in Britain were on time with 95% of delays being shorter than 15 minutes and 3 out of 1000 station stops delayed by over 30 minutes (Network Rail, n.d.). While delays are endemic in the transport system (Batley, 2007; Rezapour and Ferraro, 2021) and eliminating all of them may be impossible and not economically viable, the efforts can focus on reducing the delays that have the most negative impact on passengers or providing passengers with compensation for delays. Previous research suggests that passengers delayed by over 30 minutes are very unlikely to be satisfied with their journeys (Wittmer and Laesser, 2010; Monsuur et al., 2021). However, as noted by Transport Focus (2015), satisfaction levels tend to start dropping from the very first minute of late running.

1.2.Statement of the problem

Whilst previous research provides us with some understanding of how passenger satisfaction changes with delays, little is known about how passengers perceive delays and the effects that delay perception may have on satisfaction. Nielsen (2000) and Rezapour and Ferraro (2021) indicated that passenger perception of late running has an impact on travel behaviour and public transport suppliers can learn how to improve their services by investigating these impacts too. At the same time, many studies that focused on the valuation of changes in journey times have argued that passengers are less likely to notice smaller changes, implying that the marginal benefits of small time savings are smaller (Mackie et al., 2003; Daly et al., 2014). Despite that, Mackie et al. (2003) noted that even if passengers are unlikely to recognise minor savings, it does not automatically mean that these savings have no benefits at all. Similar arguments can be applied to delays. It can be thought that the perception of delay is an intermediate step linking the existence of delay (supply-side disruption) with the impacts on satisfaction and ultimately demand and revenue (demand-side impact).

1.2.1. How to compensate passengers for the experienced disutility?

If delays have an impact on satisfaction, but a limited impact on demand, this poses a question regarding how travellers can be compensated for the disutility related to the existence of delays. This thesis was initially motivated by the limited understanding of the role that compensating passengers for late running has on both the demand and supply side of the railways. A rail passenger delay compensation scheme aiming at improving the attractiveness of rail services and providing minimum customer service standards for delayed passengers operates in the EU and Great Britain. The scheme rules were chosen arbitrarily, and are largely homogeneous across all ticket types and journey lengths. Each year, British TOCs repay around £80m to passengers as part of the delay compensation (pre-COVID) (Gov.uk, 2020). There are two distinct features concerning the design of the rail passenger delay compensation scheme, namely the delay length threshold when passengers start receiving compensation and how the value of compensation is determined. In order to optimise the design of passenger delay compensation scheme, there is a need to better understand what levels of delays are especially inconvenient for travellers, which can be achieved by understanding delay perception and satisfaction in more detail.

1.2.2. How do delays affect passengers?

The analysis of passenger satisfaction builds on Monsuur et al. (2021) who used NRPS data to study the determinants of passenger satisfaction and the impacts of delays on the overall journey satisfaction of passengers who perceived delays. The aim of this thesis is to investigate how delays affect satisfaction with punctuality rather than overall journey satisfaction (as in Monsuur et al., 2021) of different types of passengers, both in the cases of perceived and unperceived delays. Understanding how passengers perceive delays and how satisfaction with punctuality changes with increasing delays is important and interesting in its own right. However, its usefulness in the policymaking context is limited to indicating the minimum thresholds where passengers start perceiving delays and subsequently the minimum delay thresholds that have significant impact on passenger satisfaction. The estimated thresholds can be used for determining the plausible delay distribution and performance metrics with the possibility of targeting the delays that are more likely to significantly reduce passenger satisfaction. Alternatively, these can inform the potential design of compensation mechanisms. Nevertheless, the application of the concepts of delay perception and satisfaction to economic appraisal is currently limited.

To better understand the difference between the nature of satisfaction data and the impacts of delays on satisfaction as compared to hypothetical or real choices, it is worth looking at ways in which satisfaction data can be used in deriving metrics that are typically obtained from other data sources. There is a precedent in the literature, especially covering health, labour and environmental economics, in utilising satisfaction data (particularly life satisfaction) in the context of economic valuation (e.g. Layard et al., 2008; Frey et al., 2009; Dickerson et al., 2014). This approach has, so far, not been widely used in transport. As time savings are often quantified as the largest benefit of many transport infrastructure projects, valuation of time and the impacts of reducing journey times and improving performance remain the key areas of interest for transport economists. It has been noted throughout the literature that performance improvements may often have larger benefits as compared to travel time reductions (e.g. de Jong et al., 2007). Stated preference surveys are typically used to estimate the relative valuation of delays respective to scheduled journey times (e.g. Batley and Ibáñez, 2012). However, due to the aforementioned limitations of the SP data and the different nature of the satisfaction data, it may be of interest to contrast how lateness valuation estimates vary depending on the type of data source used. Hence, this thesis also aims to explore that link by combining the methodology used in studies focusing on lateness valuation using stated preference data (e.g. Bates et al., 2001; Preston et al., 2009;

Batley and Ibáñez, 2012) and studies using satisfaction data in economic valuation (e.g. Dickerson et al., 2014) to derive reliability multipliers that are used in demand forecasting. This ensures that the concept of satisfaction is better translated to the currently used methodologies, allowing making comparisons and better translation of the results for policymaking.

1.3. Aims and objectives

The aim of this thesis is to investigate rail passengers' perceptions of delays and the consequential impacts on satisfaction. In this sense, the thesis explores the intermediate steps that link the occurrence of delay and its possible impacts on demand and revenues, as for delays to have such impacts, they need to be perceived and have a material influence on travellers' attitude and behaviours (in this case measured by reported satisfaction). The particular focus is on understanding the impacts of smaller versus larger delays on passengers, given that smaller measured delays may not be noticed or regarded as significant by travellers. The research can be helpful in designing/rethinking passenger delay compensation and devising performance strategies and targets for the British railways. This aim will be accomplished by the following objectives and the corresponding research questions:

1) Review of the currently used passenger delay compensation mechanisms (Chapter 4)

Q1: What are the costs and benefits of the currently operating passenger compensation scheme?

2) Examination of how passengers perceive delays (Chapter 6)

Q1: What are the minimum delay lengths perceivable by travellers?

Q2: How do passengers perceive the lengths of delays they experienced?

Q3: Do journey length, purpose, comfort and arrival versus departure delay impact upon how delays are perceived?

3) Assessment of the impacts of delays on passenger satisfaction (Chapter 7)

Q1: How does the probability of being satisfied with punctuality change with increasing levels of recorded delays?

Q2: What are the delay lengths detrimental to passenger satisfaction?

Q3: What are the ways in which other journey aspects, i.e. journey length and comfort affect passenger satisfaction?

Q4: Is the impact of delays on satisfaction non-linear?

4) Contrasting the concepts of delay perception and passenger satisfaction (Section 7.5)

Q1: Is there a gap between the moment a delay is perceived and starts having an impact on satisfaction?

5) Determining the relative valuation of lateness to scheduled journey time using satisfaction data (Chapter 8)

Q1: Is there any difference in the estimated trade-offs between delay and scheduled journey time based on the type of data used?

1.4.Thesis structure

This thesis follows the structure outlined below and depicted in Figure 1:

Chapter 2 Impacts of delays on passengers

This chapter summarises the theoretical models of trip scheduling as well as introduces the methods typically used to measure delay impacts. More extensive literature reviews form parts of each of the empirical chapters.

Chapter 3 Data and methodology

This chapter briefly describes the types of modelling approaches used in the thesis. However, each of the empirical chapters contains a more detailed description of the analysis undertaken.

Chapter 4 Rail passenger delay compensation scheme

This chapter evaluates the currently operating rail passenger delay compensation scheme in Great Britain. It consists of a qualitative review of scheme rules and passenger engagement, and an empirical quantitative analysis of the costs of running the scheme for different operators in Great Britain. This chapter serves as a motivation for further research conducted as part of the thesis.

Chapter 5 Data: The National Rail Passenger Survey (NRPS)

This chapter describes the main dataset used in the analysis conducted in Chapters 6-8, the National Rail Passenger Survey, its contents and limitations.

Chapter 6 Rail delays and travellers' perception of being delayed

This chapter introduces the concept of delay perception. Binary logit models are used to estimate the probability of rail travellers perceiving a delay for increasing lengths of recorded delays.

Chapter 7 Impacts of delay on travellers' satisfaction

This chapter analyses the impacts of recorded delays on reported satisfaction. Binary and ordered logistic methods are used to investigate how different levels of delays affect passenger satisfaction. The concepts of delay perception and journey satisfaction are reconciled to investigate the existence of a gap between a length of delay that is perceivable and one that has a detrimental impact on passenger satisfaction. Moreover, the potential non-linearities in the impact of delays on passengers are examined.

Chapter 8 Lateness valuation using satisfaction data

This chapter proposes an approach utilising journey satisfaction data in the estimation of an established metric – the lateness multiplier - that represents a conversion rate between a minute of delay to an equivalent of scheduled journey time.

Chapter 9 Conclusions

This chapter summarises the results of the analysis conducted in the previous chapters along its limitations. Suggestions for future work and main implications for the policymakers are also introduced in this chapter.

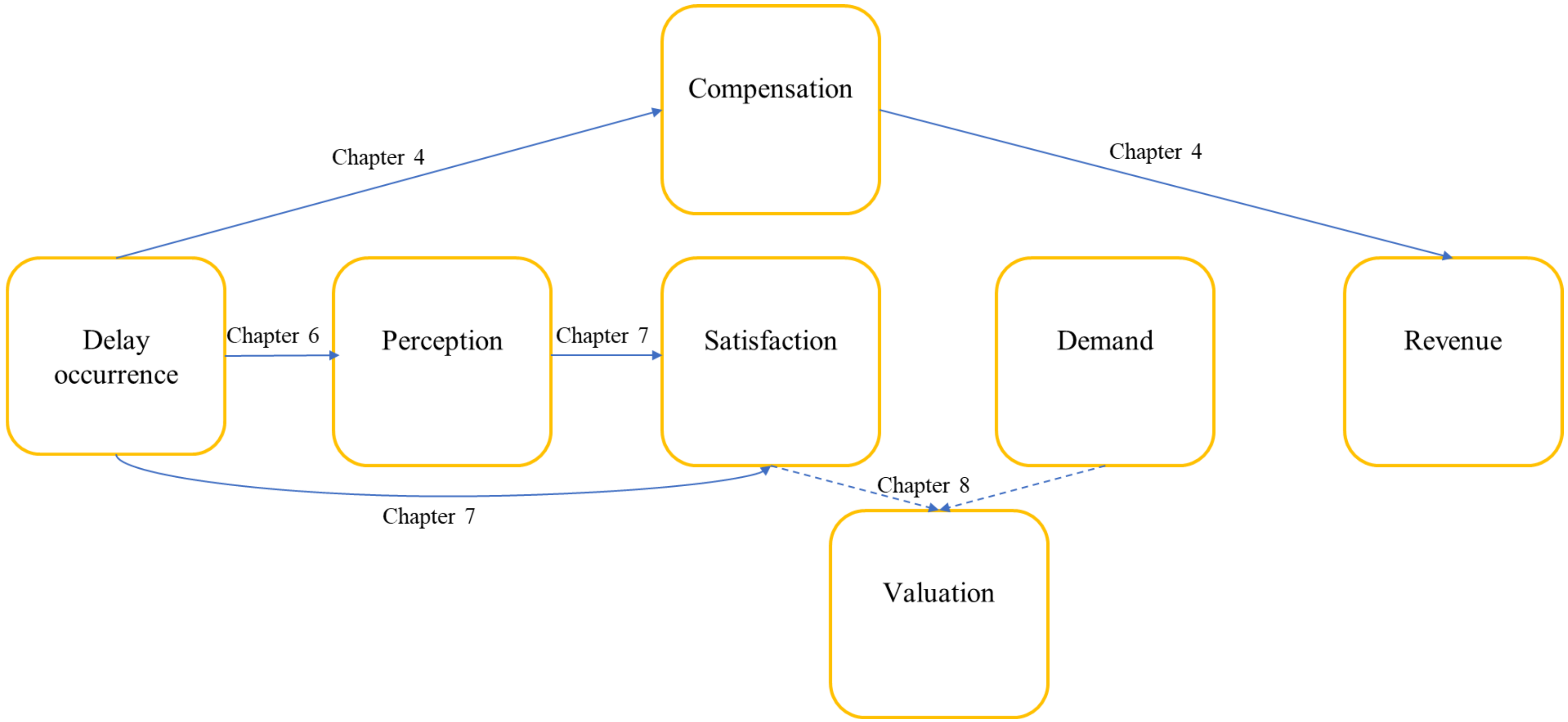


Figure 1 Thesis structure

Figure 1 depicts the chapters of the thesis and the different impacts of delays that these investigate. Chapter 4 looks at the cost of delays in terms of passenger compensation. Chapters 6 and 7 describe the immediate steps between the occurrence of delay and its impacts on demand and revenue, namely delay perception and the consequential impacts on satisfaction. The direct link between journey satisfaction and demand response is not established, however, Chapter 8 aims to bridge the two concepts by utilising satisfaction data in studying delay valuation (providing an indirect link, represented by the dotted lines).

Chapter 2
Impacts of delays on passengers

2.1. Introduction

To better understand the negative impacts that delays have on passengers, one needs to consider how travellers make travel choices and what affects them. The aim of this chapter is to provide a summary of theoretical models of trip scheduling as well as a brief review of methods typically used to measure the delay impacts.

It is generally agreed throughout the literature that reliability has an important, yet relatively small impact on travel choices (e.g. Batley, 2007). Having said that, de Jong et al. (2007) suggested that the benefits of improved reliability may often be larger than those of reduced travel times - as travellers place a large value on performance. As noted by de Jong et al. (2007), most of the research covering the value of reliability uses one of the following three approaches:

1) mean versus variance approach where unreliability is represented by the standard deviation or variance of the distribution of journey times,
2) percentiles of travel distributions where unreliability is represented by the difference between the 80th or 90th percentile of the journey times distribution and their mean or
3) scheduling models where unreliability is represented by the number of minutes a traveller departs and/or arrives early or late as compared to their Preferred Arrival Times (PATs).

The value of reliability is typically estimated from SP studies where travellers are presented with multiple-choice sets characterised by different levels of travel attributes (including the lengths of delays). Subsequently, researchers calculate the chosen metrics (e.g. mean or standard deviation of delays) and the respondents' relative valuations. While the first two approaches are more concerned about the overall performance, i.e. the distribution of delays, the latter looks at how expectations of unreliability impact trip scheduling. This approach is of particular relevance in investigating the impact of incidental lateness on passengers, which is the main focus of this thesis. The average and incidental lateness are, in fact, two distinct concepts as the impact of a given delay episode is different from the impact of general railway performance. In other words, a delay episode of 5 minutes is different from an average delay of 5 minutes. Hence, with some (deliberate) exceptions, this thesis looks at how incidental delays affect travel experiences.

2.2. Scheduling models

The basis for the trip scheduling models can be sought in the studies concerning theories of time allocation (Becker, 1965; DeSerpa, 1971) with time/money and scheduling constraints (Vickrey, 1969; Small, 1982) under uncertainty (Small, 1982; Noland and Small, 1995; von Neumann and

Morgenstern, 2007). The major difference between the private car and public transport services can be seen in the latter providing less flexibility in terms of departure and arrival times that are bound by the timetables. Hence, Bates et al. (2001) and Batley (2007) discussed the discrete nature of the departure/arrival times of public transport. Considering the work by Small (1982) and Noland and Small (1995), it is worth noting that travellers attach disutility to longer travel times, but also to late or early arrival. Hence, the concept of a 'schedule delay', defining the difference between preferred and actual arrival times. Typically, it is assumed that travellers assess all the possible travel options and choose one that maximises their utility, what in transport is equivalent to the lowest generalised cost, i.e. monetary cost and travel time (Bates et al., 2001). Travellers typically have a preferred arrival time and assign utility to each of the travel options based on how distant the scheduled arrival is from their preferred PAT (Bates et al., 2001) or PAT band (Mahmassani and Chang, 1986) - with schedule delay being a source of larger disutility for late than early arrival. Bates et al. (2001) also extended the framework described above, considering that the choice of departure time faced by travellers is made under uncertainty that is introduced by unreliability. As such, any delays or early arrivals may affect passengers' expectations of travel time and encourage the choice of earlier or later departure in the future.

A concept related to the previous discussion is that of a safety margin that defines the extra amount of time incorporated into the journeys, serving as insurance against any possible delays. This is likely to depend on both the expectations about delays (that can depend on previous experiences) and sensitivity towards late arrival (Gaver, 1968; Knight, 1974; Bates et al., 2001; Batley, 2007). Hence, it is likely that travellers are less sensitive to a delay that allows them to reach their destination within their preferred arrival time window and with the inclusion of larger safety buffers, the lengths of such delays increase. Nevertheless, the experienced delays may be larger than the included safety margins and, as noted by Bergström and Krüger (2013), in such cases, travellers may face long waits or need to make changes to their scheduled activities.

2.3. Measuring the impacts of delays

The impacts of delays are typically evaluated using:

1) market-level analysis focusing on the impact of delays on demand (e.g. Batley et al., 2011)
2) SP or RP surveys analysis of travel choices and the relative valuations of travel attributes (e.g. Batley and Ibáñez, 2012) or
3) analysis of passenger satisfaction (e.g. Monsuur et al., 2021).

With reference to journey scheduling, it has to be noted that travellers' preferences and risk aversion are likely to be heterogeneous. Hence, the benefits arising from reliability improvements are likely to have a differing impact on travellers (Batley, 2007). However, the demand analysis is aggregated at both demand and supply levels as it looks at the impacts of average delay on total demand. SP or RP surveys are also typically used to analyse the impacts of average delays (or their distribution) and their valuations. There is a large body of literature concerned with evaluating the impacts of delays on demand and valuation using SP and RP surveys. However, with these typically being concerned with the distribution of delays, a very limited number of studies looks at the impact of individual delays on passengers. Here, it would not be expected for a single delay to have a large impact on travel choices. However, such delays are a source of disutility in line with the scheduling models discussed in the previous section.

As noted in the recent paper by Wardman and Batley (2022), the implied elasticities (i.e. from SP surveys) are typically larger than these estimated from the demand data, suggesting that delays are a source of disutility, however, may not ultimately lead to changes in travel behaviour as demand studies suggest that demand is relatively inelastic with respect to performance (ATOC, 2004; Batley et al., 2011). Similarly, with SP surveys relating to hypothetical choices, there might be differences in the estimated impacts of delays on passengers based on the type of data used (i.e. SP versus satisfaction surveys).

Travellers are risk-averse and uncertainty of arrival time can cause stress and anxiety (Preston, 2008; Peer et al., 2012). Most delays are small and passengers are likely to prepare for the possibility of encountering them. However, the occurrence of longer delays that have a relatively low probability of occurring, but high impact, is more difficult to predict and prepare for (Bergström and Krüger, 2013). As eliminating all delays is practically impossible, in order to know whether it is more important to target the many small or the few very large delays, it is necessary to understand how travellers are impacted by delays of differing sizes (Bergström and Krüger, 2013).

Hence, conforming to the framework presented above, with the travel history impacting expectations about reliability, travellers schedule their journeys with respect to the preferred arrival time, also including a safety margin to their schedules. Subsequently, the experienced delay will lead to a disutility with its magnitude likely dependent on how far the actual arrival is from the preferred arrival. Subsequently, the disutility derived from a delay incident can be captured by satisfaction data where travellers evaluate their travel experience *ex-post* as shown in Figure 2. Eventually, the only way to compensate passengers for the disutility related to delays is to provide them with monetary compensation, as the time lost cannot be returned. Hence, the presented concepts ask for increased research investigating the impact of delays on passengers using the alternative data sources. It is thought that passenger reports or journey evaluation data may be a useful addition when studying the described problem (Preston, 2008).

Figure 2 Framework behind journey evaluation

Chapter 3

Data and methodology

This chapter aims to provide a summary of the data and methodologies used throughout this thesis to provide a clearer picture of the concepts introduced in each chapter. Table 1 serves as a roadmap and provides an overview of each of the chapters. Objectives, research questions and motivations associated with each of the chapters are presented alongside the methodology and data used in the analysis.

Chapter 4 provides a literature review on the impacts of passenger delay compensation as well as an econometric analysis of its costs using financial and performance data from the British franchised TOCs obtained from the Department for Transport. The remaining chapters mostly utilise logistic regression techniques to investigate how delays are perceived and the impacts they have on passenger satisfaction. These methods are applied to passenger responses from NRPS, a rail passenger survey in the UK, obtained from Transport Focus, and linked to performance data from HSP (obtained from National Rail). These data sources are described in more detail in Chapter 5.

Table 1 Summary of data and methodologies used in the thesis

Chapter	Objective (based on section 1.3)	Research question	Motivation	Methodology	Data
Chapter 4 Rail passenger delay compensation scheme	Review of the currently used passenger delay compensation mechanisms (Chapter 4)	Q1: What are the costs and benefits of the currently operating passenger compensation scheme?	Delay Repay has been introduced as a means of compensating passengers experiencing severe delays and to regulate the minimum customer service requirements for the treatment of passengers following late-running (Department for Transport, 2016). Department for Transport (2020) noted that while the proportion of passengers claiming compensation has been increasing, only 39% of surveyed passengers who experienced a delay qualifying for compensation decided to engage with the process in 2018.	Literature review on the design and operational characteristics of the scheme. Econometric analysis of the costs of the scheme and impacts on different operators.	Rail industry data sourced from the Department for Transport and ORR
Chapter 6 Rail delays and travellers' perception of being delayed	Examination of how passengers perceive delays (Chapter 6)	Q1: What are the minimum delay lengths perceivable by travellers? Q2: How do passengers perceive the lengths of delays they experienced?	Nielsen (2000) and Rezapour and Ferraro (2021) indicated that the perception of delays has an impact on travel behaviour and public transport suppliers can learn about ways in which they can improve their services by investigating it. Daly et al. (2014) suggesting that passengers are often not able to perceive small changes in travel times. Wardman and Batley (2022) and Rong et al. (2022) suggesting that research is needed to understand the differences between perceptions of late time and recorded	Binary logistic models of delay perception; analysis of the distribution of reported and recorded delay lengths	NRPS – passenger reports about delays; HSP – recorded performance

		Q3: Do journey length, purpose, comfort and arrival versus departure delay impact upon how delays are perceived?	delay lengths to better understand how delays affect passengers.		
Chapter 7 Impacts of delay on travellers' satisfaction	Assessment of the impacts of delays on passenger satisfaction (Chapter 7)	Q1: How does the probability of being satisfied with punctuality change with increasing levels of recorded delays? Q2: What are the delay lengths detrimental to passenger satisfaction? Q3: What are the ways in which other journey aspects, i.e. journey length and comfort	Suggestions that performance may not lead to large changes in demand, due to the lack of viable alternatives particularly in the short-run (Batley et al., 2011). Limited research linking recorded performance with reported satisfaction. Though, generally performance has been suggested to have a very strong impact on satisfaction (Transport Focus, 2015; Carrel et al., 2016; Gao et al., 2018; Monsuur et al., 2021). Monsuur et al. (2021) analysed the impact of recorded delays on passenger satisfaction using the NRPS dataset. However, some methodological differences are proposed. Wardman and Batley (2022) argued that proportional elasticities (i.e. based on the relative proportion of AML to GJT) better explain changes in demand than the actual delay lengths. Gao et al. (2018) proposed a cubic relationship between the difference in the experienced versus expected lengths of delays and satisfaction.	Binary and ordered logit models of passenger satisfaction (to study the impact of incidental delay on individual passenger satisfaction); fractional outcome logit regression (to study the impact of average delay on the aggregated satisfaction levels). Maximum likelihood estimation of the elasticity of marginal utility of delay; Estimation	NRPS – reported satisfaction; HSP – recorded performance

		affect passenger satisfaction? Q4: Is the impact of delays on satisfaction non-linear?	Data from satisfaction surveys has been applied to study non-linearities, for example in the context of the marginal utility of income (e.g. Layard et al., 2008).	of a cubic relationship and piecewise regression between recorded delays and reported satisfaction;	
	Contrasting the concepts of delay perception and passenger satisfaction (Section 7.5)	Q1: Is there a gap between the moment a delay is perceived and starts having an impact on satisfaction?	Same as for chapters 6 and 7	Binary logit model of delay perception and satisfaction.	NRPS – passenger reports about delays and reported satisfaction; HSP – recorded performance
Chapter 8 Lateness valuation using satisfaction data	Determining the relative valuation of lateness to scheduled journey time using satisfaction data (Chapter 8)	Q1: Is there any difference in the estimated trade-offs between delay and scheduled journey time based on the type of data used?	Monsuur et al. (2021) analysed the impact of recorded delays on passenger satisfaction using the NRPS dataset. SP surveys are typically used to estimate lateness multipliers (e.g. Bates et al., 2001; Preston et al., 2009; Batley and Ibáñez, 2012; Wardman and Batley, 2022). However, some limitations of SP survey data were highlighted in the literature (e.g. Wardman, 1988). There is a large body of literature using data from surveys on life satisfaction in economic valuation, especially in health or environmental economics (e.g. Ferrer-i-Carbonell and van Praag, 2002; Frey et al., 2009). However, such approaches have not been as widely applied in the transport contexts (with the exception of Dickerson et al., 2014).	Ordered logit model of passenger satisfaction to derive lateness multiplier.	NRPS – passenger reports about delays and reported satisfaction; HSP – recorded performance

Chapter 4
Rail passenger delay compensation scheme

4.1. Introduction

Eliminating all delays is practically impossible and probably not optimal (Batley, 2007; Rezapour and Ferraro, 2021). At the same time, previous research suggests that delays have negative impacts on passengers, but their ability to respond to worsening performance is limited (Batley et al., 2011). Following a delay, the immediate way of compensating for the additional disutility related to that is to refund the monetary value of the time lost. Against this background, and to improve the competitive position of rail, a rail passenger compensation scheme has been introduced in the UK and the EU with the aim of providing minimum customer service standards for delayed passengers.

So far, there has been very limited research empirically reviewing how the rail passenger delay compensation schemes work in practice. The amount of compensation repaid by each train operating company (TOC) as part of the scheme depends on:

1) the rules of the scheme,
2) the number of passengers eligible to claim compensation and
3) passengers' levels of engagement with the scheme.

Currently, British passengers can claim a portion of their original fare for delays of over 15 minutes with the compensation thresholds set arbitrarily. At the same time, as noted, by Wardman and Batley (2022) and Rong et al. (2022), little is known about how passengers perceive delays or what levels of delays are detrimental to their satisfaction. Hence, designing appropriate compensation mechanisms is challenging.

Assuming common speeds and equal probability of encountering a delay across the whole network, longer (and hence more expensive) journeys would be subjected to longer delays, hence TOCs operating long-distance services would be likely to see more passengers being eligible to claim compensation. This is likely combined with higher engagement rates as travellers state they are more likely to submit compensation claims for longer delays and/or more expensive journeys (Department for Transport, 2020). Hence, operators serving longer journeys may see higher claim rates due to differences in the disutility of lateness or opportunity cost of not claiming compensation. While this is likely an oversimplification, it can be hypothesised that this can lead to a differing revenue burden that the Delay Repay (DR) scheme has on the operators' revenues.

Whilst it may be easy to review the costs of the scheme by looking at the amount of compensation repaid to travellers, its potential benefits are more difficult to measure. This would require quantifying the impact of the scheme on demand (i.e. the scheme either encouraging rail travel or limiting the loss of the revenue in the future). Due to the relative complexity, limited data and its sensitivity, this chapter mainly focuses on the effects of the design and mechanics of the DR scheme on its costs. Whilst the specific focus is on the scheme operating in the United Kingdom, comparisons are drawn to the rail passenger delay compensation schemes operating within other European countries as well as a similar scheme operating in aviation.

The purpose of this chapter is to improve understanding of the role that the scheme currently has in the British railways and give recommendations and highlight research directions that might guide policymakers and regulators in the process of redesigning such a scheme in the future.

The remainder of this chapter is structured as follows:

- Section 4.2 provides a review of the British railway industry to position the DR scheme.
- Section 4.3 provides information on how the currently operating scheme works in Great Britain. Comparisons to similar schemes operating in other countries and in the airline market are subsequently presented in sections 4.4 and 4.5.
- Section 4.6 provides a qualitative analysis of passenger engagement with the scheme and a quantitative analysis of the scheme's costs.
- Section 4.7 summarises the research conducted as part of this chapter as well as provides some directions for further research and comments on the potential implications for policymaking.

4.2. Background

Delays are one of the crucial aspects of journey, affecting levels of demand, mode, route or travel time choices with significant heterogeneities in sensitivities to delays across different types of travellers (Balcombe et al., 2003; Paulley et al., 2006; Preston et al., 2009; Batley et al., 2011; Holmgren, 2013). Passengers anticipating some level of disruption based on previous experiences or due to their sensitivity to lateness usually allow some extra buffer time to their schedules as a safety margin to increase the probability of arriving to their destination within the preferred time window (Bhat and Sardesai, 2006; NEXTOR, 2010). Depending on the nature of the journey, travellers perceive 1 minute of delay as being 1 to 6.5 times worse than 1 minute of scheduled journey time (Bates et al., 2001; Preston et al., 2009; Wardman and Batley, 2014; Nagy and

Csiszár, 2015). Most of the commuter journeys are, for example, relatively short, but occur regularly, meaning that repetitive disruption can especially affect these passengers (Zahavi, 1974; Zahavi and Talvitie, 1980; Marchetti, 1994; Joly, 2004). Following worsening performance, passengers can respond by increasing their safety buffer, changing operator, mode, time of travel or decide not to travel at all, but such responses will depend on the availability of alternatives (Preston et al., 2009; Batley et al., 2011). The Delay Repay scheme has been introduced in the UK as a means of compensating passengers experiencing severe delays and to regulate the minimum customer service requirements for the treatment of passengers following late-running (Department for Transport, 2016). The scheme rules in the UK are mostly homogeneous across all ticket types and have been chosen arbitrarily with all TOCs offering passenger compensation equating to 50% of their ticket price for delays of more than 30 minutes and 100% for delays of over 1 hour. In addition, 13 TOCs also provide compensation for delays of between 15-30 minutes equating to 25% of the original ticket price.

Both abandoning the journey and spending more time travelling following a delay incur loss of social welfare. Operators may decide to compensate passengers for the resulting loss due to ethical reasons, regulation, competition or to prevent potential demand losses in the future. The only way to compensate for the increase in generalised cost following late running is to repay passengers an appropriate proportion of the fare component of the generalised cost. In the short-run, passengers can respond to lower performance by submitting compensation claims. However, if the late running occurs regularly, some passengers may likely try to find an alternative. Nevertheless, as argued by Wardman and Batley (2014), passengers will not always be able or willing to change their travel behaviour as a result of poor performance, at least in the short or medium term. Some travellers may also decide to increase their safety buffers which will, in turn, increase their generalised cost of travel – and this will not be captured by any compensation schemes.

Britain’s railway industry is characterised by vertical separation of train operation and management of rail infrastructure. Train Operating Companies (TOCs) pay track and access charges to the infrastructure manager, responsible for the maintenance of tracks and stations (Pollitt and Smith, 2002). At the same time, they are also subject to regulation and a set of performance regime mechanisms that aim to ensure safety, protect passengers’ interests and facilitate cooperation (Pollitt and Smith, 2002; Nash et al., 2013). ‘Schedule 8’ is a system involving payments between train operating companies (TOCs) and the infrastructure manager (IM) based on the marginal revenue effect of delays where affected parties compensate each other for the effects of late running on their long-term revenue (Network Rail, n.d.; Wardman and Batley, 2014; ORR, 2016; Steer, 2018). The aim of the scheme (in principle financially neutral) is to incentivise TOCs and IM to invest money in preventing delays and reduce costs related to financial risks caused by delays (Network Rail, 2012; Wardman and Batley, 2014; Steer, 2018). Schedule 8 payments are currently unrelated to passenger compensation that TOCs are required to pay to delayed passengers as part of franchise agreements (ORR, 2014), though it has been proposed that the two schemes become more interlinked (ORR, 2021). As argued by ORR (2014), whilst both schemes reflect on performance, they serve different roles, with Schedule 8 relating to compensation and incentive arrangements between TOCs and IM and the DR scheme serving as a means of compensating passengers for delays.

The DR scheme might be seen as an additional cost of delays with TOCs having little incentive to encourage passenger engagement unless the costs of the compensation scheme are seen as a prevention against future revenue loss. Without extensive regulation or competition, the TOCs may be incentivised to make the process of engaging with the scheme more difficult (costly) for passengers. Passengers may value the existence of a compensation scheme, but it remains difficult to estimate the impact the existence of the scheme has on demand or compare the benefits of the scheme (increased revenue) with its costs (compensation paid). In fact, little is known about the benefits of the scheme and the impact of the scheme's design on its costs, which are driven by passengers' eligibility to claim compensation and the levels of passenger engagement with the claiming process. Whilst eligibility is exogenous, being determined by the scheme rules and driven by performance, the impact of varying engagement on the scheme's costs has not been empirically tested. It can be expected that engagement levels (propensity to claim compensation) may differ between passengers due to:

- differences in sensitivities to the experienced levels of lateness as is the case with journey time or fare elasticities of demand (Bates et al., 2001; Preston et al., 2009; Wardman and Batley, 2014) and
- as a result of the expected costs and benefits of claiming compensation (opportunity cost of not claiming compensation).

Whilst this has generally been confirmed by passengers who state the length of the delay and ticket price as major factors determining their engagement (Department for Transport, 2020), it is not known how these differences impact the compensation values and revenues of British TOCs.

4.3. Rail passenger delay compensation scheme in Great Britain

The rail passenger delay compensation scheme introduced in the UK is a more generous version of the scheme adopted by the European Parliament in 2007. The EU directive

1371/2007 details the level of compensation passengers are entitled to claim for severe delays - 25% of the ticket price paid for delays between 60 and 119 minutes and 50% for delays of over 2 hours with minimal compensation that can be set at up to 4 euros (European Commission, 2007).

The original compensation scheme offered on the British railways, referred to as the Passengers' Charter, was based on arrangements stipulated by the National Rail Conditions of Travel. Under these rules, travellers were eligible to receive 50% of the single ticket price for delays of over 60 minutes. This scheme has now been replaced by DR.

All of the British franchised TOCs are required to provide compensation for passengers affected by delays of over 30 minutes with a number of TOCs voluntarily paying compensation for delays of over 15 minutes as detailed in Table 2.

Table 2 Delay repay scheme rules in the United Kingdom

Delay	Compensation
15-29 minutes	25% of single tickets (only selected TOCs decided to implement that voluntarily)
30-59 minutes	50% of single tickets
60-119 minutes	100% of single tickets
>120 minutes	100% of return journey

It is noted that the rules shown in Table 2 apply to non-seasonal tickets. However, season ticket holders are also eligible for compensation for individual incidences of delays. In such cases, the compensation value for a single delay incidence is determined on the assumption of 464 journeys made on an annual ticket. Hence, a single journey ticket price equivalent is calculated by dividing the total ticket cost by 464. Similar principles apply to other types of seasonal tickets. Prior to the introduction of Delay Repay, under the Passengers' Charter, season ticket holders were eligible for compensation in the form of season ticket renewal discounts. Travellers were able to receive a 5% or 10% discount when their TOC did not meet performance targets.

Both the guidelines provided by the European Commission and the UK's implementation of the scheme have somewhat been arbitrary with no documented economic research into the effects that late running has on passengers and the value passengers place on such a scheme contrasted with the costs of the scheme.

In the UK, the total amount of compensation paid by TOCs in 2009/10 equalled £2.9m compared to £89.4m paid in 2019/20 as shown in Figure 3. It must, however, be noted that the values are not directly comparable as the current scheme is very different from how it was functioning earlier. First of all, under the Passengers' Charter the compensation values also included discounts for seasonal ticket holders. Moreover, while Delay Repay had already been introduced in 2007, it has taken almost 15 years for all the franchised TOCs to adopt it with Chiltern being the last TOC to join (Haylen, 2019). The levels of compensation between 2009 and 2015 were visibly lower what can likely be attributed to the Passengers' Charter offering more limited compensation as well as lower popularity of the scheme and/or higher costs of submitting claims due to limited automation.

Since 2016/17, the compensation payments have been oscillating between £74m-£84m (not adjusted for inflation). In 2019, a number of TOCs also introduced a 'DR15' scheme where passengers are eligible to claim 25% of their ticket price following a delay of over 15 minutes. At the time of conducting this research (2019), only two TOCs – Cross Country and LNER had not introduced it (Gov.uk, 2020). The introduction of the DR15 scheme in 2019 contributed to an additional £5.2m repaid to passengers, though the timing of the introduction of the scheme differed across the TOCs and with only one (not full) year of data (Gov.uk, 2020), analysing the impact of the DR15 scheme is currently not in the scope of this work. Generally, DR compensation averaged around 1% of ticket revenue, ranging between 0.1% to almost 3% for individual TOCs in 2019.

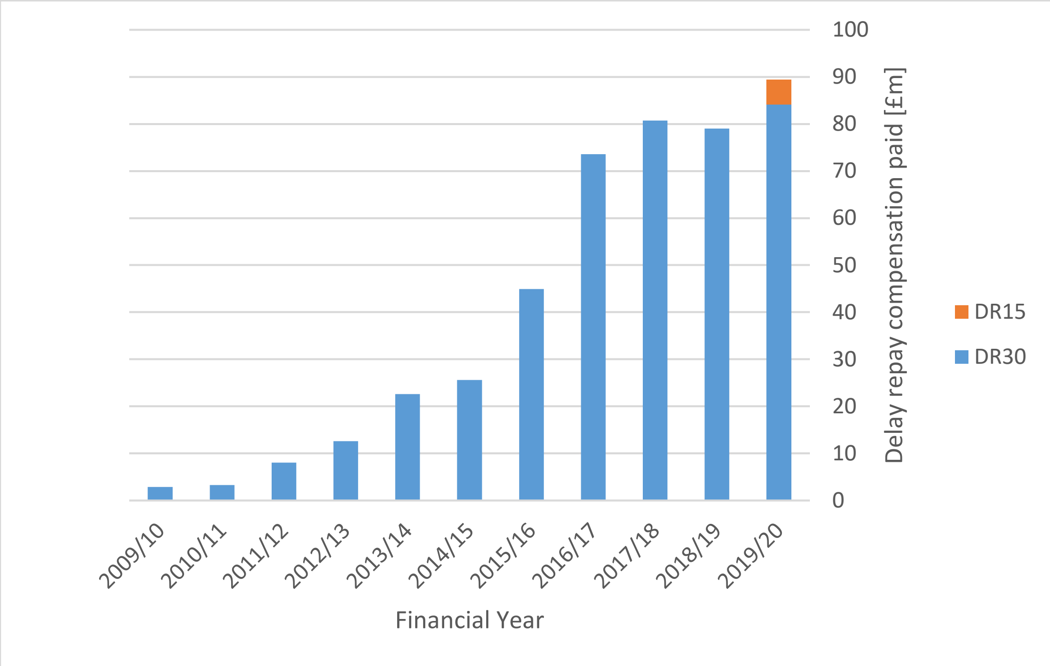


Figure 3 Delay compensation paid to rail passengers in the UK between 2009-2019 (Source: Gov.uk, 2020)

To better understand the design of the DR compensation scheme, Figure 4 shows the levels of compensation available for journeys with scheduled in-vehicle time of 15, 30 and 60 minutes for increasing actual in-vehicle times. The compensation thresholds are not dynamic, i.e. whether the delay is 35 minutes or 45 minutes does not affect the level of compensation available. Similarly, the compensation thresholds are the same for all journeys, regardless of their lengths. Three types of journeys were chosen for a more detailed investigation to depict the mechanics of the scheme – a shorter, commuter-type journey, a medium-length journey and an extremely long journey (as shown in Table 3). While these are not necessarily representative of average passenger experiences, the comparisons allow us to better understand the relationship between journey length, delay length, fares and compensation available to passengers. The comparisons in Figure 5 show compensation per 1% increase in journey time (because of delay) as the metric of interest, since it allows us to track the relationship between the relative change in journey time due to delay and the monetary compensation. While it has to be recognised that journey length, distance and ticket prices are unlikely to be perfectly correlated, assuming that longer journeys are typically also more expensive, the value of compensation per 1% increase in journey time is larger for longer journeys (i.e. Aberdeen to Penzance). The respective compensation per 1% of journey time added is smaller for shorter journeys as these are less expensive and a 30-minute delay represents a much larger relative increase in journey time. Nevertheless, the compensation per 1% of journey time added initially decreases with delay length increasing from 30 to 59 minutes. After 60 minutes of delay, the eligible compensation equals full fare and compensation per 1% of journey time added is at the same level as for the 30-minute delay and subsequently starts decreasing again, highlighting the impact of the non-dynamic compensation thresholds.

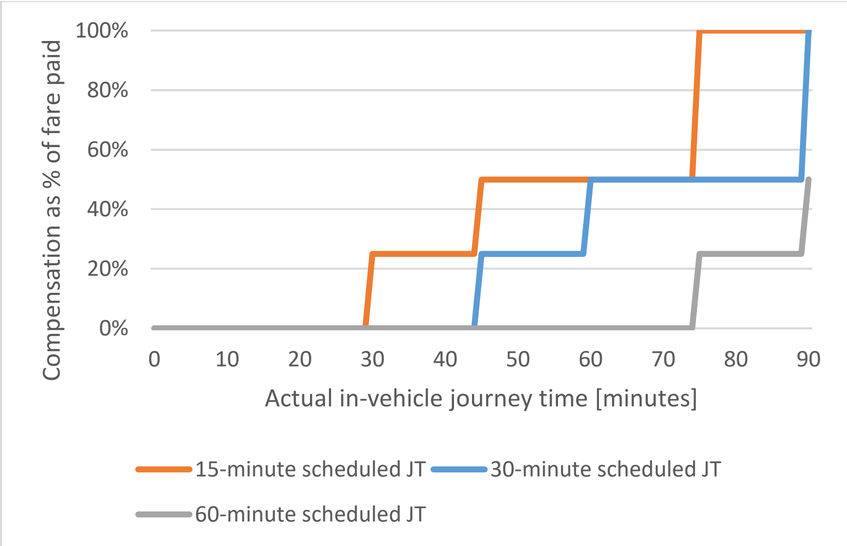


Figure 4 Relationship between actual journey time and compensation

Table 3 Summary of journeys selected for further investigation

Origin	Destination	Journey time (minutes)	Price
Ilkley	Leeds	28	£5.40
Nottingham	London	113	£74.00
Aberdeen	Penzance	843	£252.25

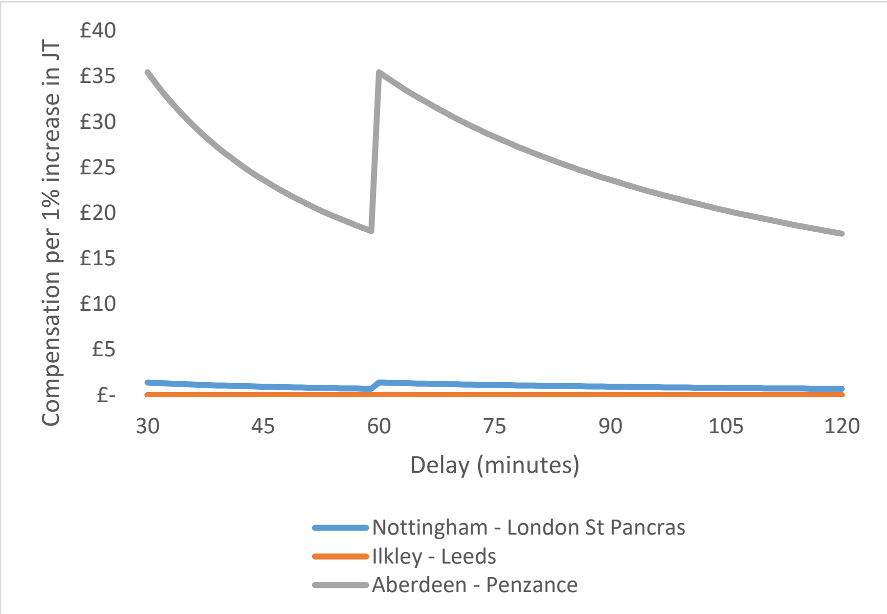


Figure 5 Relationship between delay length and compensation per 1% increase in journey time for selected journeys

4.4. Comparison with passenger delay compensation schemes operating in other European countries

The previous section described the rail passenger delay compensation scheme in Great Britain. As noted before, the basis for the introduction of the scheme dates back to European Commission Directive 1371/2007. The European Commission recommended a scheme allowing passengers to receive 25% of the ticket price paid for delays between 60 and 119 minutes and 50% for delays of over 2 hours, with minimal compensation set at up to 4 euros (European Commission, 2007). In principle, these rules apply to services across the EU, however, it is noted that there might be some local differences related to how the scheme works in practice. Urban, suburban and regional services may be exempt from these rules, but operators are not exempt from paying compensation in the case of *force majeure* as specified in the judgment related to case C-509/11 against ÖBB-Personenverkehr AG. Since then, several European countries or operators introduced the scheme or its modified version. The scheme rules frequently change, but a comparison across countries and operators may give some indication of the differences in the recent scheme rules across the continent. Table 4 presents the rules for selected countries and operators that are interesting case studies and offer a scheme that is different from the scheme based on EU directive 1371/2007.

Some countries, like Great Britain, the Netherlands or Belgium modified the scheme, providing more compensation and for shorter delays. In most cases, the scheme rules do not differentiate between different types of services and journey lengths. However, in the case of Spanish operator Renfe, compensation thresholds differ between different types of services, with high-speed rail travellers being eligible for higher relative compensation and for shorter delays. In case of Czechia, an open-access operator, RegioJet, operates a more generous scheme than the incumbent České dráhy. This may be aimed at improving their competitive position in the market. Moreover, their compensation scheme has quite complicated rules as the proportion of fare that passengers can receive changes with journey lengths and is based on whether the responsibility for the delay is attributed to the operator.

Outside of the EU and UK, there are also examples of delay compensation with Swiss National Railways offering the EU-style compensation, Canadian Via Rail offering travel credits for longer delays and Korean National Railroad providing compensation of 12.5% of fare for delays of over 20 minutes. In some countries, delay compensation is offered for the more expensive, high-speed services, e.g. Tejas Express in India or high-speed services in Taiwan.

Table 4 Rail passenger delay compensation scheme rules for selected EU countries and operators

Delay (minutes)	EU directive 1371/2007 (Germany, Czechia, Poland)	Great Britain	Spain (Medium distance)	Spain (AVE high-speed)	The Netherlands	Belgium	RegioJet (<1.5 h, operator's fault)	RegioJet (<1.5 h, not operator's fault)
15-29	-	25%*	25%	50%	-	-	-	-
30-44	-	50%	50%	100%	50%	-	50%	25%
45-59	-	50%	50%		50%	-	50%	25%
60-89	25%	100%	100%		100%	100%	100%	25%
90-119	25%							100%
>120	50%							

*selected train operators

Percentages represent the portion of fare passengers can reclaim following a delay of specified length

Source: Operators' websites as of 2022

Comparison of the compensation values repaid in various countries is difficult. First of all, the data availability in most cases is very limited. In the case of the UK, the compensation data is available from the regulator’s website. An attempt was made to review websites and reports conducted by local governments, operators and regulators to search for comparable data from other European countries. Compensation data for some countries may exist, but due to language barriers, the ability to search for it was limited. However, in a communication with the Polish regulator (UTK), as of November 2019, it was confirmed that the regulator did not collect any data on delay compensation paid to passengers. Nevertheless, some aggregated data about compensation volumes or numbers of submitted claims was successfully found for Spain, Germany and Austria. Another characteristic that limits the possible comparisons is related to the differences in the scheme rules. While these are quantifiable, there might also be some technical differences related to how passengers submit compensation claims, how compensation is paid as well as some country-specific characteristics, e.g. travel costs and journey times that can also affect the compensation levels and are more difficult to quantify. Hence, direct comparisons may be impractical, but the analysis of the compensation repaid in other countries can provide some useful context.

As discussed before, in the UK, the total amount of compensation paid by TOCs in the recent (pre-COVID) years was around £80m per annum. In Germany, 2.7m rail passengers were compensated in 2018 with the total compensation reaching €53.6m (The Local, 2019). This represents just over half of the total volume of compensation repaid in the UK in the corresponding time period. In Spain, 256 claims were submitted per 100,000 passengers on high-speed/long-distance journeys and 105 on medium-distance journeys in 2019 (Renfe, 2019). The number of compensation claims submitted for rail delays in Austria increased from 4,800 in 2009 to 35,000 in 2013 (a 7-fold increase) with the value of compensation paid changing from €275,000 in 2010 to €360,000 in 2012 and over €600,000 in 2015 (APF, 2015; 2018). Whilst these headline data provide some valuable insights, they are aggregated and not directly comparable between the countries.

4.5. Comparison between passenger delay compensation scheme operating in railways and airline market

Whilst the main focus of this chapter is on the delay passenger compensation scheme operating in railways, it is worth commenting on a similar scheme operating in the airline market. There are some unique characteristics of the two modes of transport that are likely to justify the differences in passenger delay compensation schemes between airlines and railways, but in principle, the two schemes are similar. This section aims to review the airline compensation scheme and compare it to the scheme operating in railways.

In 2004, the European Commission adopted 261/2004 legislation, establishing common rules regarding compensation and assistance to passengers whose flights are delayed or cancelled within the Bloc (European Commission, 2004). Current regulations, i.e. 261/2004 legislation (European Commission, 2004) set the compensation levels based on route length, length of delay, and its cause. Passengers delayed due to external factors (that are outside of the airline's control) are not eligible for any compensation. However, the airline is obliged to take care of passengers and ensure reasonable arrangements are made to provide the service. For eligible air passengers affected by a delay, the compensation is not linked to the ticket price (as is the case with railways). Hence, there is a possibility of compensation being larger than the ticket price, which is often criticised by the airlines (ERA, 2019). Levels of compensation set out by 261/2004 regulation (European Commission, 2004) are detailed in Table 5.

Table 5 Airline compensation scheme rules in the EU (European Commission, 2004)

	Length (km)	Delay	Compensation (euros)
Short-haul	<1500	>3h	250
Medium-haul	1500-3500	>3h	400
Long-haul	>3500	>4h	600

The definition of the 'external factors' has been a cause of multiple disputes with several cases taken to court (Europe Economics, 2019). The disputes were with regards to whether delay causes such as airport security and airline staff strikes or faults in the functioning of an aircraft can be regarded as extraordinary circumstances or not. Examples of court cases are C-549/07 Wallentin-Hermann and B2/2013/3277/CCRTF Huzar (Europe Economics, 2019). The difficulties in establishing clear criteria for what can be classified as the disputable extraordinary circumstances and the high levels of compensation available (as the compensation, unlike in rail, is not linked to the ticket price) have caused the emergence of third parties. These act as intermediaries between passengers and airlines in the process of submitting compensation claims and usually operate a no-win, no-fee model (Europe Economics, 2019). These companies use social media platforms for marketing to increase passenger awareness and encourage them to use their services. Kenny Jacobs, chief marketing officer for Ryanair (one of the leading low-cost airlines in Europe), claimed in an interview that for every £10m of compensation, £4-£5m is directed to the third parties (The Independent, 2018).

With the level of compensation not being linked to ticket price and the arrival of low-cost airlines which often offer tickets priced at even less than 10 euros, the amount of compensation passengers could be entitled to in many cases is several times larger than the original fare. Hence, especially passengers travelling on low-cost airlines may benefit more from the compensation while the low-cost airlines may be more affected by the levels of compensation they have to pay. The European Regions Airline Association conducted a review study of the current EU261 regulation and surveyed employees of selected airlines to gain insights into the industry's perspective (ERA, 2019). The report focused on smaller, regional airlines and argued that the compensation represents large portions of the revenue margins and leads to reductions in the range of services offered by the operators. For the airlines involved in this study, the compensation paid to passengers increased by over 300% between 2016 and 2019 with the average compensation representing 300% of an average ticket price. The report recommended some changes to the regulations, including exoneration of PSO routes from the scheme, limiting compensation to a proportion of airfare, extending the delay threshold from 3 to 5 hours and acknowledging the knock-on effect of extraordinary circumstances on the whole daily flight programme (ERA, 2019). Finally, it was argued that the application of rail compensation scheme rules (in the form recommended by the European Commission) would lead to almost trebling of the number of eligible passengers and halving of the scheme's costs (ERA, 2019). Moreover, the results of an anonymous survey conducted among airline employees suggested that most of the employees felt that the current regulation has a negative impact on safety.

The current provision of delay compensation differs between the railway and airline industries. There might be some differences in the characteristics of air and rail travel that provide reasoning behind the different models of delay compensation for the two modes. For example, in the European context, it is highly unlikely that an experienced rail delay stretches to more than a couple of hours due to a relatively dense and small network, high frequency of departures, availability of alternative modes in the event of disruption or ability of train operating companies to provide a replacement relatively quickly. Flights, on the other hand, tend to be less frequent, meaning that in case of a delay it may be more difficult to find an alternative quickly (i.e. flights cannot always be as easily replaced by a bus or taxi journey, as is the case in railways). Taking all these into account, it is more likely for air passengers to be affected by lengthier delays than for rail passengers. As a result, air passengers tend to include large safety margins in their journeys to increase the probability of arriving at their destination before their preferred arrival time (NEXTOR, 2010).

The summary of major characteristics of the rail and air delay repay schemes in the UK is provided in Table 6. Currently, rail passengers can be compensated up to the maximum of fare paid, whereas, airline passengers can only get compensation for extreme delays (only caused by airlines) with compensation not being constrained by the ticket price. Similarly, as in the case of the rail compensation scheme, the airline compensation scheme does not offer any compensation for passengers below the selected thresholds that, unlike in rail, vary and depend on the category of journey lengths. Therefore, compensation is only available for passengers affected by larger delays, usually only when the actual journey time is double or triple the scheduled journey time. It is, therefore, important to investigate whether there is a threshold beyond which passengers should get compensated more than they paid for the ticket, both in the rail and airline contexts and at what levels of delays rail and air passengers should start being compensated. Whilst such considerations are not in the scope of this study, it is recommended that policymakers investigate these in the future.

Table 6 Comparison of rail and airline compensation schemes

Characteristic	Airline	Rail
Cause of delay	Only eligible for compensation if the delay is caused by the airline	Eligibility irrespective of who/what caused the delay in the case of Great Britain
Value of compensation	Not linked to the fare paid and can be larger than the ticket price	Represents a proportion of ticket price
Availability of information to the operating company	Airlines generally have near perfect information on the number of passengers, their journeys and characteristics	TOCs usually have more limited information about the number of passengers affected by delays

In recent years, a large increase in demand for air travel has been fuelled by a rapid expansion of low-cost carriers (LCC) and the increase in demand generally had a negative impact on performance (Dobruszkes, 2006; Pratt and Schuckert, 2018). Bhadra (2009) suggested that passengers travelling on discounted fares are less likely to submit complaints. This may be a result of LCC passengers accepting lower quality in return for lower travel costs (O'Connell and Williams, 2005; Bhadra, 2009; Chiou and Chen, 2010). However, O'Connell and Williams (2005) suggested that there are also significant differences in the characteristics of passengers choosing traditional and low-cost airlines. Nevertheless, Bhadra (2009) argued that this process can be seen as a mutually beneficial exchange that can be disrupted by extensive regulation. Forbes (2008) argued that passengers complain if the quality is worse than what they expected. Therefore, another reason why LCC passengers are less likely to complain may be due to an expectation of lower quality. An example application of the low-cost carrier phenomenon to railways could be an introduction of special low-fare tickets with limited eligibility for compensation (reducing revenue risk to TOCs). Studying this phenomenon further could also help understand the differences in passenger engagement in the claiming process based on different journey or operator types. However, as argued by the Department for Transport (2016), competition in the airline sector has resulted in larger heterogeneity of customer care and regulation may still be needed to set the minimum customer care standards. Whilst competition is greater in airlines and the extent of existence of low-cost carriers in airlines cannot be translated to rail, Stead et al. (2019) suggested that in the GB rail, open access operators have been typically scoring better in terms of passenger satisfaction despite worse performance in terms of punctuality, what could be due to lower expectations. Department for Transport (2016) suggested that it is not clear whether increased competition would be able to address any potential issues with the quality of customer service in railways, highlighting the need for including compensation schemes as part of franchise agreements with TOCs. On the other hand, Czechia is an example of a country where a new entrant RegioJet offers a more generous compensation scheme than the incumbent (RegioJet, n.d.; České dráhy, 2016). Nevertheless, it is currently difficult to understand the motivation behind the provision of a different scheme by the open-access operator and its possible impacts on the respective market shares of the operators.

4.6. Analysis of the British 'Delay Repay'

Section 4.3 introduced the rules of the rail passenger delay compensation scheme operating in Great Britain. Sections 4.4 and 4.5 summarised the compensatory mechanisms for travellers operating in other European countries and the airline market.

Noting that the specific focus of this thesis is the railway market in Great Britain, Table 2 introduced the rules for the British rail compensation scheme. The analysis conducted in the following sections aims to investigate how this scheme works in practice, how its rules affect its operations and passenger engagement, and the impacts it has on train operator revenues. Understanding the impacts of the scheme on passengers and operators and contrasting its costs and benefits is currently not in the scope of the work conducted as part of this chapter as the focus remains on the costs of the currently operating scheme.

With regard to the benefits, it may be expected that the scheme has an impact on demand through:

- encouraging more demand due to travellers knowing that if they are delayed, they can be compensated or
- limiting the demand loss related to worsening performance.

It is, however, unlikely for an econometric analysis of the relationship between the compensation values and ticket sales to provide any insights into the size of any of the two aforementioned effects as they are, in fact, difficult to be observed. A more appropriate line of analysis would possibly include studies using stated preference surveys. These could provide some insights into travellers' willingness-to-pay to be protected by a delay compensation scheme or its impacts on travellers' stated mode choices. As this chapter focuses on the costs of the DR scheme, it is appropriate to consider factors affecting compensation levels.

In principle, the compensation levels depend on:

- the number of passengers eligible to claim compensation,
- the proportion of eligible passengers who submit compensation claims and
- the value of compensation for an eligible passenger.

Having ignored the heterogeneity in ticket prices and experienced delays for different journeys, the total compensation can be represented as:

$$Total\ Compensation = \kappa \times \omega \times Demand \times \zeta \times Average\ Revenue \tag{1}$$

where:

$Demand$ and $Average\ Revenue$ are exogenous

κ represents the proportion of eligible passengers who claimed compensation, i.e. the engagement rate

ω represents the proportion of passengers ($Demand$) eligible to claim compensation and depends on the scheme rules

ζ represents the proportion of fare available for compensation which is specified by the scheme rules (i.e. Table 2).

Whilst the scheme rules are pre-defined and directly affect both ω and ζ, both parameters are also likely to be affected by performance levels. At the same time, κ is determined by passengers' levels of engagement with the claiming process. Under the full automation of the scheme, i.e. where all delayed passengers receive compensation automatically, this would be equal to 1. This leads to a discussion about the two main drivers of compensation levels, namely eligibility (total compensation passengers could have claimed) depending on scheme rules and engagement (proportion of passengers that decided to claim compensation) depending on how many eligible passengers submit claims, i.e.

$$Compensation = f(Eligibility \times Engagement) \qquad (2)$$

4.6.1. Passenger engagement with the scheme

It remains difficult to estimate the number of passengers affected by a given length of delay due to a very large number of ticket types, large number of station stops, different origins and destinations as well as varying delays at different station stops. In most countries, performance of a rail network is measured with a focus on the supply side of delays, for example looking at the proportion of trains that arrive to the destinations within a given margin of delay (Rietveld et al., 2001; Preston et al., 2009). However, this is not informative of the proportion of passengers affected by different levels of delays. This is due to each train being treated equally and not weighted by demand. Therefore, from the passenger DR perspective, the focus of monitoring performance should be on the demand side, looking at the number of passengers affected by given lengths of delays (Preston et al., 2009; Transport Focus, 2015). Nevertheless, the delay length measured by train arrival does not necessarily represent the final delay for passengers as they can be affected by congestion at stations, crowding on board (Preston et al., 2009) and missed connections.

The number of claims and value of compensation were increasing pre-COVID, which could potentially be caused by two reasons, i.e.

- increased demand and more delays, translating to more passengers being eligible to claim or
- reduction in the costs of submitting a claim, resulting from making the claiming process easier.

In recent years, there has been a lot of interest from the regulatory bodies, the public and TOCs regarding the levels of passenger engagement with the DR scheme. Department for Transport (2020) noted that while the proportion of passengers claiming compensation has been increasing, only 39% of surveyed passengers who experienced a delay qualifying for compensation decided to engage with the process in 2018. There is evidence that passengers affected by lengthier delays and/or travelling with more expensive tickets are more likely to engage with the scheme (Department for Transport, 2019). Most of the eligible passengers who knew about the scheme but decided not to claim cited low expected compensation compared to the costs of engagement as the main reasons for choosing not to submit claims (Department for Transport, 2020). The marginal propensity to claim compensation may be non-constant as passengers quoted length of delay and ticket price as two major characteristics motivating their attitude towards the scheme with estimated claim rates ranging from 22% on Transport for Wales to 64% on LNER (Department for Transport, 2020). To better understand the drivers of engagement, the costs and benefits of applying for compensation have to be understood in more detail.

In recent years, efforts have been made to facilitate the process of claiming compensation through the introduction of automatic repayments, more available information about the scheme for customers and/or online repayment systems (Europe Economics, 2019). However, in most cases, the claim submission process is relatively time-consuming and often requires providing a photo of the ticket used, personal and journey details with passengers having to create an account for each of the operators separately. Selected TOCs now offer automated compensation for season or advanced tickets, however, the impact of automation has not been understood very well so far. The process usually, in fact, requires some initial effort plus a "one-click" process to submit claims, reducing, but not totally removing the marginal costs of submission. Fully automating the process, while reducing administrative costs of the scheme, would require the usage of smart ticketing as the current ticketing system does not allow for accurate tracking of passenger journeys due to the existence of anytime tickets not matched to just one service. As suggested by Railway Technology (2020), this also leads to fraudulent claims.

To better understand what influences the proportion of eligible passengers applying for compensation, the costs and benefits of applying for compensation have to be understood in more detail. As suggested by Europe Economics (2019), the costs of applying for compensation can be divided into three steps of submitting a claim:

1) becoming aware of the 'Delay Repay' scheme,
2) gathering information about the eligibility and claiming process and
3) claim submission.

Mainstream economics assumes that consumers' decisions are rational and bounded by their personal preferences and available information (Europe Economics, 2019), hence assuming that travellers are rational, they only submit a claim if the value of compensation is larger than the expected costs of the claim submission. The total costs of engaging with the scheme can differ for each of the aforementioned three steps of submitting claims, as outlined in Table 7.

Table 7 Costs related to each of the steps of the claiming process (Europe Economics, 2019)

Steps of the claiming process	What influences the costs
1. Becoming aware	General availability of information about the scheme at stations, train operating companies' websites and social media. Provision of information about the experienced disruption.
2. Gathering information about eligibility criteria and the claim submission process	Availability of information about eligibility, value of potential compensation and how to submit a claim.
3. Claim submission	Difficulty of submitting the first and subsequent claims (i.e. creating an account for the first submission and then filling in application forms).

The first two steps can be thought of as the initial costs that are only incurred at the first instance of delay where a delayed passenger decides on whether or not to engage with the scheme. The costs for the third step have to be incurred for each of the claim submissions unless there is a fully automated compensation system. This mechanism is conceptually similar to switching bank accounts or energy providers with the potential for benefits after investing time and effort to engage with the process (as discussed in Klemperer, 1995; Wilson and Price, 2005; The Social Market Foundation, 2015; Europe Economics, 2019).

When thinking about engaging with the scheme, a passenger chooses whether or not to claim compensation based on the disutility created by the delay (this will be further explored as part of this thesis) and the expected benefits (minus costs of submitting claims) which include the monetary compensation but are not limited to that form of compensation. Europe Economics (2019) argue that there are behavioural biases that influence the decision on whether to submit claims that are not based on the assumed rationality, i.e.

1) **Behavioural bias** may increase passenger engagement with the scheme as delays can be perceived as unfair. Receiving compensation (even if small related to the time spent submitting a claim) may be perceived as a benefit per se.
2) Especially the more frequent travellers perceive a possibility of encountering more delays in the future. Therefore, they may see incurring the initial costs as an investment for the accumulation of benefits in the future, which is called a **projection bias**.
3) Even if the benefits to costs ratio would indicate on claiming to be rational (i.e. large compensation for a small amount of time spent on sending a claim), the **default behaviour**, however, is not to engage with the scheme regardless of the expected benefits. Evidence from the pension 'auto-enrollment' schemes suggests that the default effects have large impacts on behavioural choices (Europe Economics, 2019; Hardcastle, 2012; Leicester et al., 2012).

Hence, it is of interest to investigate how journey and passenger characteristics affect levels of compensation repaid by different TOCs. These can include journey purpose, length, fare, delay length and/or journey comfort. Similarly, passenger characteristics such as age, income, education level and access to internet may also affect claiming behaviour. However, analysing impacts of passenger characteristics is more difficult and would require conducting passenger surveys.

It is crucial to test the hypothesis that the marginal propensity to claim compensation increases with fares and/or delay lengths as suggested by Europe Economics (2019) and Department for Transport (2020). As longer delays and higher ticket prices mean higher eligible compensation, claiming compensation for longer delays is naturally more worthy of investing time and effort. Moreover, some types of passengers may be more sensitive to delays as shown by research into delay elasticities (Wardman and Batley, 2014; Wardman and Toner, 2020; Wardman and Batley, 2022) and higher engagement levels may also result from the aforementioned behavioural biases.

Summarising, the total amount of compensation repaid to passengers depends on how many passengers are eligible to claim compensation and the percentage of eligible passengers that submitted claims. Eligibility, which can be understood as the total compensation passengers could have claimed, depends on performance, and fare levels - and is determined by the scheme rules, which are predefined. Engagement, on the other hand, is expected to increase with delay length and ticket price. The following section will aim to quantify these effects.

4.6.2. Quantitative analysis of the impact of 'Delay Repay' on operator revenues

To explore the impacts of the DR scheme on the operators' revenues, data on performance, compensation, operation, and revenues were obtained for the British franchised TOCs from the regulator's website. At the time of collating the data, most of it was available annually for at least 5 years between 2015 and 2019. TOCs and the regulator as of now (2023) publish monthly data on the number of claims and compensation volumes. However, caution would be needed in comparing pre-COVID data used as part of this analysis with the data from the COVID and post-COVID times as these periods may be characterised by structural changes in railway usage.

As compensation represents a percentage of ticket price, it is natural that (assuming the same levels of performance and engagement) TOCs characterised by more demand repay more compensation, with average compensation increasing in line with fares. As shown in Figure 6, between 2015 and 2020, the average compensation was below 20 pence per passenger journey for most TOCs. There are, however, two TOCs characterised by an average compensation of between 20 to 100 pence per passenger journey, namely Avanti and LNER, both mainly operating long-distance services and employing the same version of DR as most of the other operators. Further analysis aims to analyse the impact of TOC characteristics on the scheme costs.

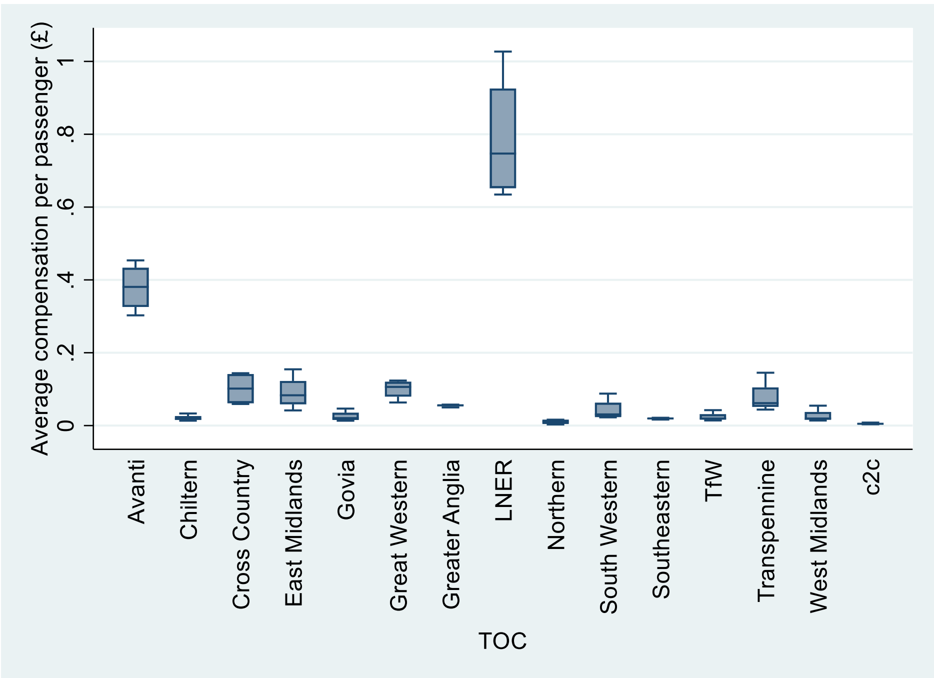


Figure 6 Boxplot of compensation per passenger journey for years 2015-2020

The impact of average fare on total compensation can be divided into four distinct effects that were identified based on the work by Europe Economics (2019):

1) **Engagement effect**: increased claim rates for journeys with higher fares due to anticipation of larger benefits and differences in passenger types or journey lengths leading to heterogeneity in the disutility of lateness.

2) **Eligibility effect**: if longer journeys (in terms of distance) are more delayed (i.e. journey length and delay length are correlated), the incidence of longer delays that qualify for compensation is larger among TOCs operating such services.

3) **Revenue effect**: if the TOC's average fare is higher, compensation per passenger journey will also be higher as compensation represents a proportion of ticket price.

4) **Fare increase effect**: changes in compensation per passenger between years could be explained by changes in average fare resulting from inflation.

With fare increases being marginal compared to the differences in average fare between different TOCs, the revenue effect is a natural phenomenon, resulting from the scheme's design and passengers' behaviour does not have any impact on this. For direct comparisons and to enable inferences about engagement levels, it is necessary to control for the revenue effect. Therefore, compensation to ticket revenue ratio (CRR) is used as the primary variable of interest, i.e.

$$CRR = \frac{Total\ Compensation}{Total\ Revenue} \times 100\% \qquad (3)$$

Whilst the focus so far has been on compensation, it is necessary to look at the relationships between all of the aforementioned variables. The compensation, represented by CRR, is a function of eligibility and engagement as shown in Figure 7. With eligibility being mostly affected by the pre-defined scheme rules and operator's performance, the engagement is suggested to depend on both fare levels and performance while both are likely to be correlated with journey time or distance (though likely not perfectly). The levels of engagement will also depend on the journey type and passenger characteristics, though when using aggregate (i.e. operator-level) data, it is not possible to disentangle these effects.

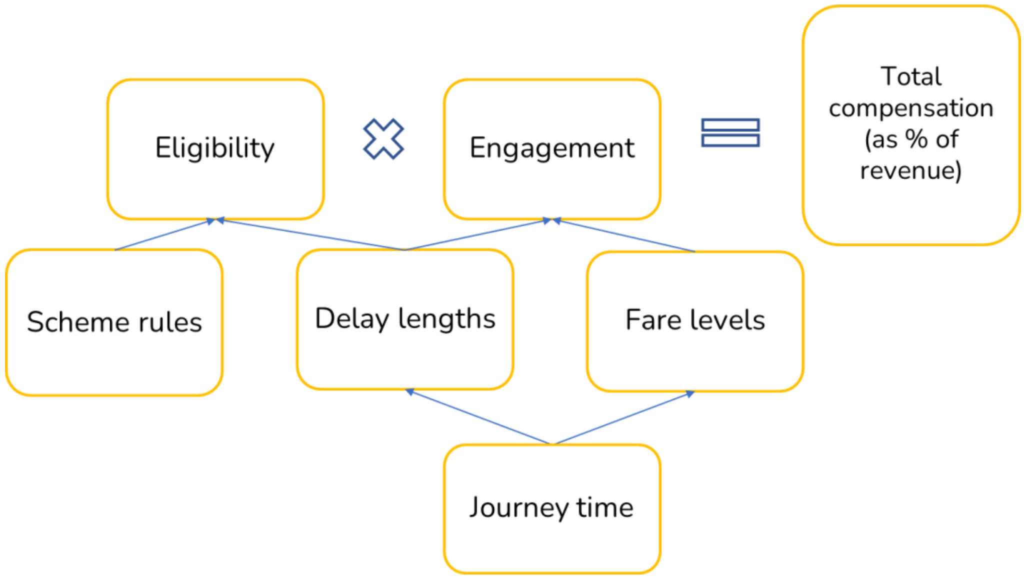


Figure 7 Schematic of the relationships between variables

Two existing metrics reported by ORR can be particularly useful in studying the impact of performance on compensation levels:

1) Average Passenger Lateness (APL)

APL reported by ORR represents the estimated length of delay an average passenger on the British rail network is subjected to. This can be thought of as the mean of passenger delay distribution, but the compensation scheme only depends on the number of passengers affected by the more severe (and relatively more uncommon) delays. Therefore, this will depend more on the skewness of the delay distribution, rather than its mean. It can, however, be expected that average lateness generally increases with the increased incidence of longer delays.

2) The proportion of station stops delayed by over 15 minutes

As passengers are eligible to claim compensation only for severe delays, the distribution of delays is important in determining this and the metric reporting the proportion of larger delays is useful in understanding the shape of lateness distribution. However, it is focused on the supply side of delays, weighting delays by station stops. The two statistics are, as expected, highly correlated (r=0.95). It can be expected that with more station stops being severely delayed, an average passenger experiences a longer delay.

Table 8 summarises the distribution of values for the selected variables related to performance, operational characteristics and compensation for the British franchised TOCs.

Table 8 Summary statistics (comparison between TOCs)

Variable	Minimum	Maximum	Mean	Standard deviation
APL (min)	0.8	10.5	3.78	1.95
Average Fare (£)	3.0	36.8	10.7	9.8
Average Journey Length (km)	24.5	260.6	71.8	66.3
Claims per 1,000 passengers*	0.5	21.6	5.3	5.2
Compensation per passenger journey (pence)	0.3	102.7	11.8	20.7
CRR (%)	0.1	2.8	0.7	0.5
Station stops delayed by over 15 minutes (%)	0.3	8.7	2.8	2.1

*2018 and 2019 only

The TOCs were divided into three categories based on average journey length, representing short (up to 50 km), medium (50-100 km) and long-distance journeys (over 100 km) as shown in Table 9. This categorisation acts as a proxy for differences in passenger and journey characteristics that may affect levels of engagement for different journey lengths. Additional categorisation was based on whether the TOC operates within South East of England where London is a major attraction as similar segmentations have been used in, for example, fare elasticities recommended by PDFH (ATOC, 2004) (highlighting the potential for differences in engagement).

Table 9 Summary statistics for selected TOCs

TOC	TOC type	LSE TOC	Average length (km)	Average fare (£)	APL (min)	CRR (%)
Avanti	Long	0	197.2	31.2	7.36	1.32
Chiltern	Medium	1	55.6	8.1	2.08	0.22
CrossCountry	Medium	0	91.5	13.6	5.83	1.02
EastMidlands	Medium	0	88.6	14.7	4.03	1.05
Govia	Short	1	26.8	4.6	2.71	0.38
GreatWestern	Medium	1	61.5	10.4	3.55	1.12
GreaterAnglia	Short	1	45.5	7.6	2.55	0.74
LNER	Long	0	260	36.8	8.13	2.79
Northern	Short	0	26.8	3.1	4.02	0.51

SouthWestern	Short	1	28.0	4.8	3.22	0.49
Southeastern	Short	1	25.8	4.7	2.19	0.41
TfW	Short	0	36.8	4.5	3.66	0.93
Transpennine	Medium	0	72.0	9.0	8.43	1.61
WestMidlands	Short	1	39.6	4.9	4.62	1.12
c2c	Short	1	25.4	3.7	0.80	0.01

As the pricing model naturally suggests, longer journeys are usually more expensive with average fare per kilometre of journey length ranging from 11.6 to 18.2 pence in the investigated sample. It is noted that the opposite may be true in the cases of slower versus faster services, however, due to data aggregation, it is reasonable to assume that overall TOCs characterised by longer average journeys are characterised by larger average revenue. As expected, average journey length and average fare are characterised by an almost perfect positive correlation (r=0.99) as shown in Table 10. Average journey length is positively correlated with average passenger lateness (r=0.78), meaning that on average longer journeys are characterised by longer delays, possibly representing a smaller percentage increase in journey time. This could be due to the fact that while the total delay minutes usually increase with journey time, the marginal delay decreases with journey length due to some possible differences in journey characteristics, scheduling or capacity utilisation and demand (Armstrong and Preston, 2017; Yap and Cats, 2021). Taking all this into account, on average passengers travelling on more expensive services will usually be subjected to a longer delay overall resulting in a smaller percentage increase in journey time. This, in turn, means that while a higher percentage of passengers on the more expensive journeys will be eligible to claim compensation, higher claim rates can be expected due to longer delays and more expensive tickets as was also suggested by surveyed passengers (Department for Transport, 2020). This is likely to have an impact on the proportion of ticket revenue repaid by different TOCs.

Figure 8 shows a box plot of the number of claims received by each of the TOCs in each of the categories per 100,000 passenger journeys made in 2019. It can be immediately seen that, while the ranges are similar for short (2-700 claims per 100,000 passengers) and medium-distance (85-820 claims per 100,000 passengers) TOCs with a slightly larger median value for medium-distance TOCs, all long-distance TOCs saw a larger number of claims per 100,000 passenger journeys in 2019 (825-3400 claims per 100,000 passengers). Here, it is worth reminding that in Spain, the equivalent number of claims submitted was between 100 and 250 per 100,000 passengers, suggesting that the UK figures are typically higher.

Table 10 Correlation matrix

	(1)	(2)	(3)	(4)	(5)
(1) Compensation to revenue ratio	1.00				
(2) Average passenger lateness	0.88	1.00			
(3) % stops delayed by over 15 min	0.85	0.94	1.00		
(4) Average fare	0.81	0.75	0.84	1.00	
(5) Average journey length	0.82	0.78	0.86	0.99	1.00

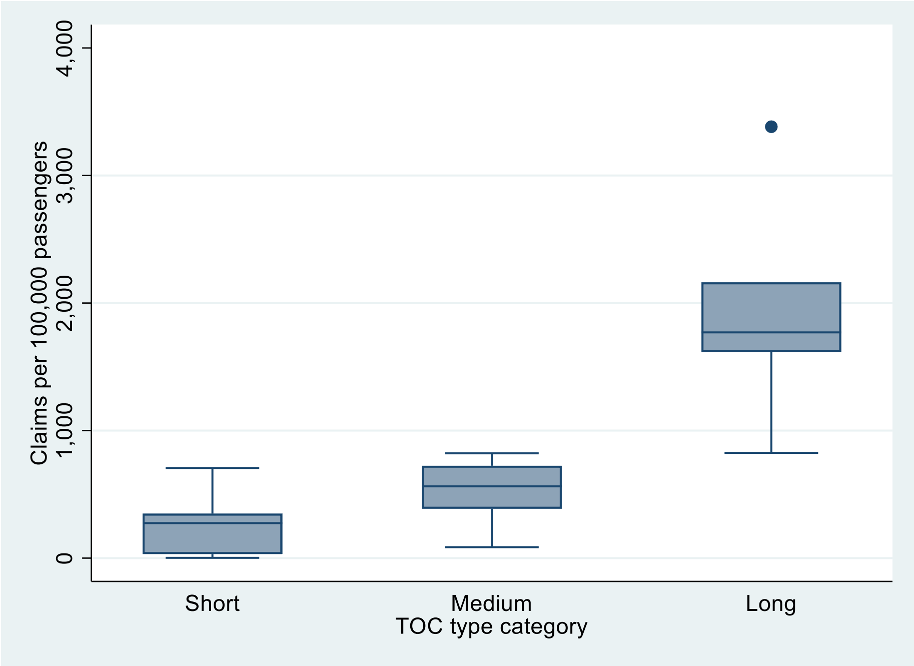


Figure 8 Boxplot of the number of claims per 100,000 passengers

The Kruskal-Wallis (Kruskal & Wallis, 1952) test was used to assess whether the differences in the within-categories distributions are statistically significant. The null hypothesis of the Kruskal-Wallis test is that the distribution of claims per 100,000 passenger journeys is similar across the TOC categories. The data were ranked from the lowest to the highest value of claims per 100,000 journeys as shown in Figure 9 for 22 TOCs where data on the number of claims was available.

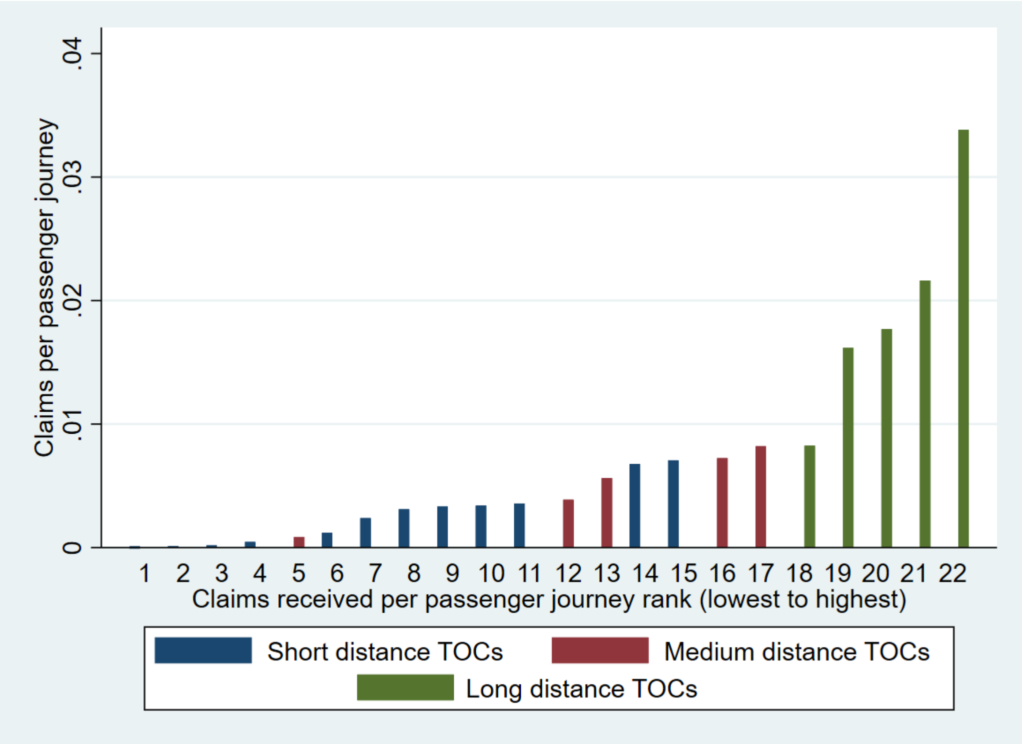


Figure 9 Ranking of claims per passenger journey by TOC type

While the null hypothesis of identical distributions was rejected (p=0.001), Dunn's post hoc test (Dunn, 1964) for each pair of categories was used to test whether the probability of observing a random value of claims per passenger journey in the first group being larger than in the second group is equal to 0.5 under null hypothesis. The null hypothesis was rejected for short-long distance pairs (p=0.0004) but failed to be rejected for short-medium and medium-long distance pairs (p-values of 0.070 and 0.053). While this provides some evidence for the fact that longer journeys attract more compensation claims, it is now of interest to assess the extent to which this affects operator revenues.

At the time of collecting the data, only two full years of data were available on the number of claims submitted to each of the TOCs, limiting the ability to analyse the impacts of eligibility and engagement on claim rates. Moreover, the claim rates per 100,000 journeys do not provide any information about the revenue impact. For these reasons, the main focus of this study remains on a more robust variable, namely the aforementioned CRR.

An econometric model was constructed to test the impact of performance levels and TOC characteristics on the compensation payments made to passengers as part of the DR scheme. Data for 4 years were used starting in 2016/17 as this is the first year where compensation payments became to be directly comparable between TOCs. The payments made as part of the DR15 were excluded, as previously discussed. It was assumed that the scheme rules are homogeneous across all the TOCs while remembering that increasing automation of the scheme through a one-click claiming process on advanced or seasonal tickets may have generally reduced the costs of submission throughout the years. To make comparisons between TOCs possible, the compensation to revenue ratio (shown in Figure 10) was used as the variable representing the scheme's revenue burden, where eligibility is determined by performance (represented by APL and proportion of stops delayed by over 15 minutes) and engagement is determined by both performance and fares (represented by APL and average fare).

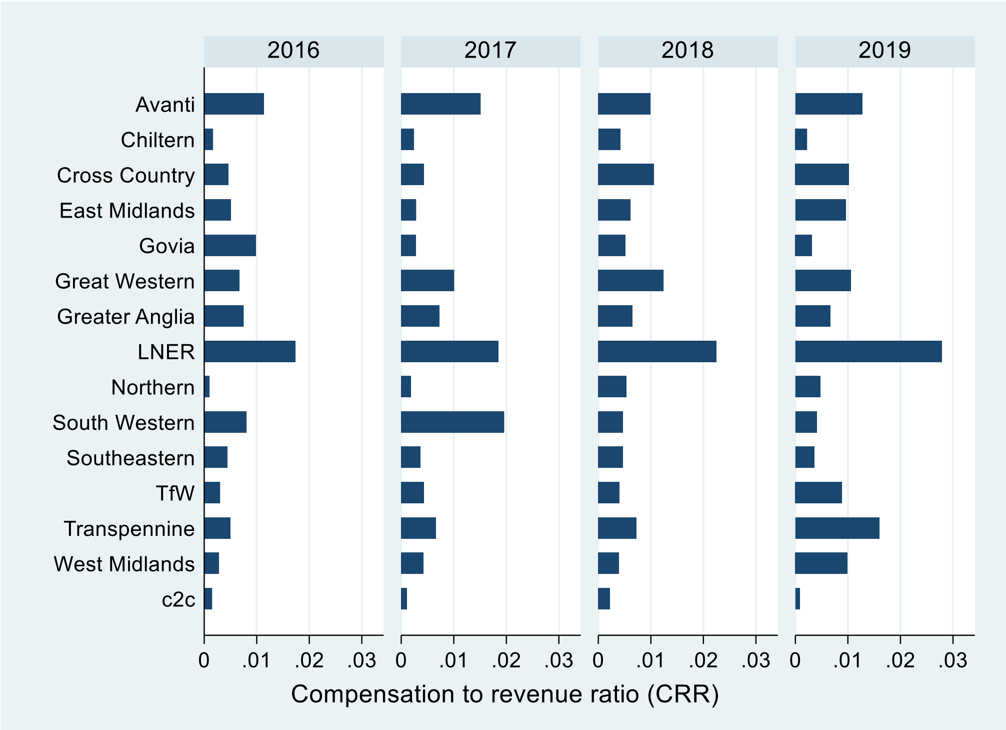


Figure 10 Compensation to fare ratio between 2016/17 and 2019/20

A more detailed inspection was conducted to better understand any possible differences in the compensation schemes offered by the 15 TOCs in the 4 analysed years. The following observations were made:

- Chiltern, for example, was offering a different version of the scheme, making direct comparisons with other TOCs impossible. The DR was only launched for Chiltern in 2022 as prior to this, travellers could only claim for delays of over 30 minutes when disruption was within operator's control or for all delays of more than 60 minutes (Global Railway Review, 2022).
- In addition to the DR payments, Govia repaid £2.2m in 2016 and £12m in 2017 to Southern season ticket holders for extraordinary disruption in 2016.
- Great Western Railway monthly and annual season ticket holders were still being offered seasonal ticket discounts rather than DR, but the impact of this on the total

compensation payments was deemed to be limited (Great Western Railway, 2019; Gov.uk, 2020).

- Transport for Wales only introduced the scheme in 2018 and SouthWestern in 2017.
- A large increase in compensation payments by LNER in 2019 cannot be fully explained by a similar increase in APL. To mitigate that, the additional dummy representing 2019 LNER will be introduced to the model as a sensitivity test.

Taking all these into consideration, Chiltern, South Western and Transport for Wales were excluded from further analysis.

An OLS model was constructed to test the impacts of both eligibility and engagement on the revenue burden (represented as compensation to revenue ratio ($CRR_{i,t}$)) of the scheme for the selected 12 British TOCs for the years 2016-2019. The equation below presents the initial specification of the model (OLS1 in Table 12):

$$CRR_{i,t} = \beta_0 + \beta_1 APL_{i,t} + \beta_2 Fare_{i,t} \tag{4}$$

where:

i : each of the TOCs in the sample

t : year

$CRR_{i,t}$: compensation to revenue ratio

$APL_{i,t}$: average passenger lateness

$Fare_{i,t}$: average fare

As shown in the equation above, the first model only includes two explanatory variables, namely the average fare and APL. APL increases both eligibility (i.e. only longer delays qualify for compensation) and engagement (as previous research suggests that engagement generally increases with delay lengths), thus the APL coefficient is a proxy for the combined effect that increased eligibility and engagement have on the compensation levels. Average fare refers to the additional effect that the increased fare (and thus journey length as both are highly correlated) has on engagement levels. Hence, β_1 corresponds to the additional engagement resulting from lateness disutility whilst β_2 corresponds to the increased engagement due to the opportunity cost of not claiming compensation.

Subsequently, the model was extended with Table 11 providing more detailed information about the additional variables used in the modelling. These include dummy variables representing TOC characteristics and time trends. Moreover, the variable describing the distribution of delays (D15) is introduced to the model and as part of sensitivity testing with average fare replaced with long distance TOC dummy variable.

Table 11 Summary of variables used in the modelling

Variable	Type	Expected impact	Comments	Included in OLS models
CRR	Continuous	Dependent variable	n/a	1-4
APL	Continuous	Positive	Represents the combined eligibility and engagement impact of performance.	1-4
D15	Continuous	Positive	Percentage of station stops delayed by over 15 minutes represents the combined eligibility and engagement impact of performance, focusing on the performance distribution.	2-4
Fare	Continuous	Positive	Represents the impact of additional engagement related to claiming compensation for more expensive tickets (opportunity cost of not claiming compensation).	1, 3-4
Long distance	Categorical, binary	Positive	Replaces average fare to represent the impact of additional engagement on long-distance journeys, which can be expected due to higher prices; included for sensitivity testing.	2
LSE	Categorical, binary	Positive or neutral	Using aggregate TOC level data, it is not possible to apply flow segmentation, typically used in rail demand research (Institute for Transport Studies et al., 2016). However, it is possible to test if passengers using TOCs operating in the South East claim more or less compensation, which may be due to higher sensitivity to lateness (increasing engagement with the scheme) or higher	2-4

			incomes (with lower marginal utility of income resulting in lower engagement levels).	
LNER 2019	Categorical, binary	Positive	Takes into account a relatively large change in CRR between 2018 and 2019 for LNER not accompanied by a large change in performance levels.	2-4
Year	Categorical, binary	Positive or neutral	Tests the impacts of potential increased automation or knowledge about the scheme with time or any potential year-related effects.	2-4

The OLS econometric model was run to test the impact of increasing lateness and fares on the revenue burden of the scheme on different types of TOCs through increased eligibility and engagement as a cross-sectional model in its simpler form (model OLS1 in Table 12), with addition of the aforementioned controls (OLS2 to OLS4 in Table 12) as well as the OLS2 version of the model rerun as a random effects panel model (RE in Table 12).

Random effects were found to be non-significant in the Breusch-Pagan Lagrange multiplier (Breusch & Pagan, 1979) (p=0.13), suggesting that a simple OLS model is run. A model with TOC-specific fixed effects is not included as it is believed that any differences in the scheme related to the claiming processes may be year, not TOC, specific.

The results in Table 12, suggest that for each £1m ticket revenue, each 1 minute of average lateness costs TOCs around £2000 in compensation. Long distance TOCs, at the same time, repay an additional £4500 (or £2000 for each £10 of average fare) while London and South East operators repay an additional £1700-£2300. LNER in 2019 repaid more than expected by the lateness levels, suggesting that an additional £9630 was repaid for each £1m revenue. The time trend was not statistically significant, suggesting that overall, the changes in the scheme have not had any significant impact on claim engagement.

It is noted that inclusion of the additional variables increases the R^2 value, as the model is able to capture the larger portion of the variation in the CRR values. Moreover, the simpler model (i.e. OLS1) suffered from non-normally distributed errors (Shapiro–Wilk test for normal data (Shapiro & Wilk, 1965), p-value of 0.0388). This improved with the addition of controls. The likelihood-ratio test (Wilks, 1938) was also run to test whether adding more predictors significantly improves fit (between OLS1 and OLS4, p-value<0.0001). In conclusion, the extended version of the model (i.e. OLS4) may be better-suited for modelling the studied relationship as the additional variables are able to capture the additional effects that average fare and APL are not able to capture on their own.

Table 12 Modelling results

	OLS1	OLS2	OLS3	OLS4	RE
APL	.0018***	.0024***	.0021***	.0022***	.0024***
	(.0003)	(.0005)	(.0004)	(.0004)	(.0005)
D15		-.0444	-.0378	-.0408	-.0559
		(.0545)	(.0476)	(.0517)	(.062)
Long distance		.0045***		.0003	.0049***
		(.0013)		(.0022)	(.0016)
LSE		.0017**	.0023***	.0023**	.0015
		(.0008)	(.0007)	(.0008)	(.0011)
LNER 2019		.0096***	.0090***	.0090***	.0096***
		(.0021)	(.0020)	(.0020)	(.0020)
2017		-.0005	-.0004	-.0004	-.0005
		(.0008)	(.0007)	(.0007)	(.0007)
2018		-.0001	.0001	.0001	.0000
		(.0008)	(.0007)	(.0008)	(.0007)
2019		.0005	.0007	.007	.0005
		(.0008)	(.0008)	(.0008)	(.0007)
Fare	.0002**		.0002***	.0002*	
	(.0001)		(.0000)	(.0001)	
Constant	-.0017*	-.0025**	-.0038***	-.0038**	-.0022*
	(.0008)	(.0010)	(.0009)	(.0011)	(.0012)
N	48	48	48	48	48
R-squared	.83	.89	.91	.92	.91

Standard errors in parentheses
**** p<.01, ** p<.05, * p<.1*

As previously discussed, APL increases with journey length and, thus, with average fare. Table 13 shows the average fare and APL averaged for short, medium and long-distance TOCs. These are then used to compute the average effects of performance on eligibility and engagement and the additional engagement effect of average fare using the different models estimated in Table 12. On average, the effect of APL on compensation increases from 0.5% of revenue on short-distance TOCs to 1.3% for long-distance TOCs. The additional engagement related to the increased opportunity cost of not claiming compensation translates to 0.1% of revenue for short-distance TOCs to 0.7% for long-distance TOCs. Both effects combined lead to an increasing burden of the scheme for TOCs operating longer journeys. Therefore, an average long-distance TOC would typically repay around 1.6% of their revenue compared to 0.7% and 0.3% for medium and short-distance TOCs respectively. Considering the fact that journey length and average passenger lateness are correlated, if long-distance TOCs repay a larger proportion of revenue, they repay more than what would be suggested by their performance levels. If this is the case, it might be needed to find an economic or regulatory explanation and reasoning for this discrepancy.

Table 13 Impacts of eligibility and engagement on the predicted costs of the DR scheme

	Short	**Medium**	**Long**
Averages based on the sample			
APL	2.72	4.63	7.01
Fare	4.75	11.80	33.61
Modelled impact of eligibility and additional engagement on CRR (OLS 1)*			
Eligibility	0.48%	0.82%	1.24%
Additional engagement	0.08%	0.21%	0.60%
Modelled impact of eligibility and additional engagement on CRR (OLS 3)*			
Eligibility	0.53%	0.85%	1.23%
Additional engagement	0.10%	0.25%	0.72%
Modelled impact of eligibility and additional engagement on CRR (OLS 4)*			
Eligibility	0.53%	0.85%	1.26%
Additional engagement	0.10%	0.24%	0.68%

Predicted CRR using average sample values (95% CI in brackets)			
Model 1	0.400% (0.31%, 0.49%)	0.862% (0.79%, 0.94%)	1.669% (1.51%, 1.83%)
Model 3	0.252% (0.12%, 0.38%)	0.726% (0.64%, 0.81%)	1.575% (1.44%, 1.71%)
Model 4	0.258% (0.11%, 0.41%)	0.723% (0.63%, 0.82%)	1.578% (1.44%, 1.71%)

*****This was computed by first setting the average fare and APL to zero and computing the corresponding CRR based on the mean values for the other variables. Subsequently, the eligibility effect was computed as the difference in the predicted CRR when setting APL to sample averages (holding fare at 0). The additional engagement effect was then computed as the difference in the predicted CRR when setting both APL and average fare at sample averages.

4.7.Conclusions

This chapter aimed to review the rail passenger delay compensation scheme currently operating in Great Britain. This was achieved by:

- providing a review of the scheme rules,
- drawing comparisons to similar schemes operating in other countries and for other modes and
- a qualitative and quantitative analysis of the relationship between performance and fares on the revenue burden of the scheme on different TOCs.

This was done to improve the understanding of the role that the scheme currently has in British railways and give some recommendations and research directions that might guide policymakers and regulators in the process of redesigning such a scheme in the future. The analysis of the scheme's benefits was currently out of scope due to limited data availability and complexity in capturing the effect that the scheme has on passengers and demand.

Rail passenger delay compensation schemes have been introduced in the EU and GB to protect the rights of delayed passengers. The scheme rules differ between the EU countries and GB, but the economic rationale behind the schemes is similar. The focus on the scheme operating within Great Britain results from a lack of suitable data for other European countries. The compensation levels were compared between British franchised TOCs to better understand the impact of the scheme on the revenues of different types of train operators. Approximately £80m was repaid to passengers every year (pre-COVID) as part of the DR scheme in GB, with TOCs typically repaying between 0.1% and 3% of ticket revenues. Whilst the scheme rules are homogeneous (i.e. the proportion of ticket price that passengers are eligible to claim back does not change with journey or delay lengths), longer journeys are typically characterised by longer delays. At the same time, longer journeys are also typically more expensive. This naturally affects the number of passengers eligible to claim compensation. This is further amplified by the marginal propensity to claiming compensation increasing with delay lengths and ticket prices, which, in turn, affects TOCs' ticket revenues. The fact that passengers are more likely to claim compensation for more expensive journeys and longer delays may be due to non-constant marginal disutility of lateness impacting differences in journey satisfaction or higher opportunity costs of not claiming compensation for more expensive journeys.

While more research is needed to understand the differences in engagement rates and possible reasons for their existence, this work provides additional evidence that the propensity to claiming compensation increases with delay lengths and ticket prices. The differences in eligibility (increasing with delay lengths) and engagement (increasing with ticket price and delay lengths) lead to significant differences in the scheme's burden for different TOCs. Other things being equal, each additional minute of APL increases the proportion of ticket revenue repaid to passengers as part of the scheme by 0.2%. As a result of engagement levels increasing with the ticket price, for the same levels of performance, TOCs repay an additional 0.2% of their ticket revenue for each £10 of the average fare. This suggests a larger financial impact of the scheme on longer-distance operators.

There are two immediate areas that would benefit from further research. First of all, more detailed data on compensation complemented by detailed ticket sales data would allow analysing the differences in eligibility and engagement at an OD pair level. It is reasonable to expect that being able to control for any OD-specific differences may further increase understanding of how the scheme works in practice. This study serves as a motivation for the regulators to require the TOCs to collect and publish more detailed data on compensation (and especially so in other European countries where only very limited data is available). This could enable further research into passenger engagement with the claiming process. Furthermore, it is recommended that a full-scale study be conducted to analyse the impacts of the scheme on passengers, revenues and, ultimately demand to contrast the scheme's benefits with its costs. It is thought that there is a potential to conduct studies utilising journey satisfaction or stated preference surveys to understand how the prospects of receiving compensation for delays affect respondents' choices or passenger satisfaction.

While this study provided evidence for the increased cost of the scheme for long-distance operators, it does not necessarily imply that the current scheme is suboptimal and needs to be changed. There might be reasons for having one fit-for-all set of rules that are easier to understand for passengers as well as operate from the administrative point of view. Nevertheless, there are examples of operators that offer a more complex version of the scheme where rules change based on journey or delay types, namely Spanish Renfe and Czech RegioJet. If the regulators aim to increase engagement with the scheme, apart from automating the process, the claiming process could also be centralised, allowing passengers to claim compensation for all journeys from the same (central) portal as was also suggested by the Williams-Shapps Plan for Rail (Department for Transport, 2021). This would reduce the initial costs of claiming, as passengers would not need to register a separate account to claim from an operator they had not claimed from before.

If a comparison of the costs and benefits of the scheme leads to a conclusion that the current design is suboptimal, it is necessary to base the design of the scheme on research concerning the impact of delays on passengers. It can particularly be useful to establish:

- the lengths of delays that are detrimental to passenger satisfaction,
- whether the negative impacts of delays vary by journey lengths and/or types,
- the potential non-linearities in the impacts of delays related to the impact of smaller versus larger delays.

Finally, while compensation currently only accounts for a small percentage of TOCs' revenues, greater automation of the scheme could contribute to increasing compensation payments (leading to TOCs repaying a larger portion of their revenues), highlighting the need for further research. It also needs to be noted that the analysis focused on the pre-COVID period and it might be beneficial to consider how the DR scheme has been impacted by the COVID pandemic.

Motivated by the analysis of the DR scheme conducted as part of this chapter, the remaining chapters of this thesis focus on analysing the impacts of delays on travellers what could provide an additional source of information and guidance should an operator or regulator (in GB or other European country) decide to redesign passenger compensation scheme. Hence, the remaining chapters will focus on delay perception and the consequential impacts on satisfaction with the following chapter introducing the dataset used in the subsequent analysis.

Chapter 5
Data: The National Rail Passenger Survey (NRPS)

5.1. Introduction

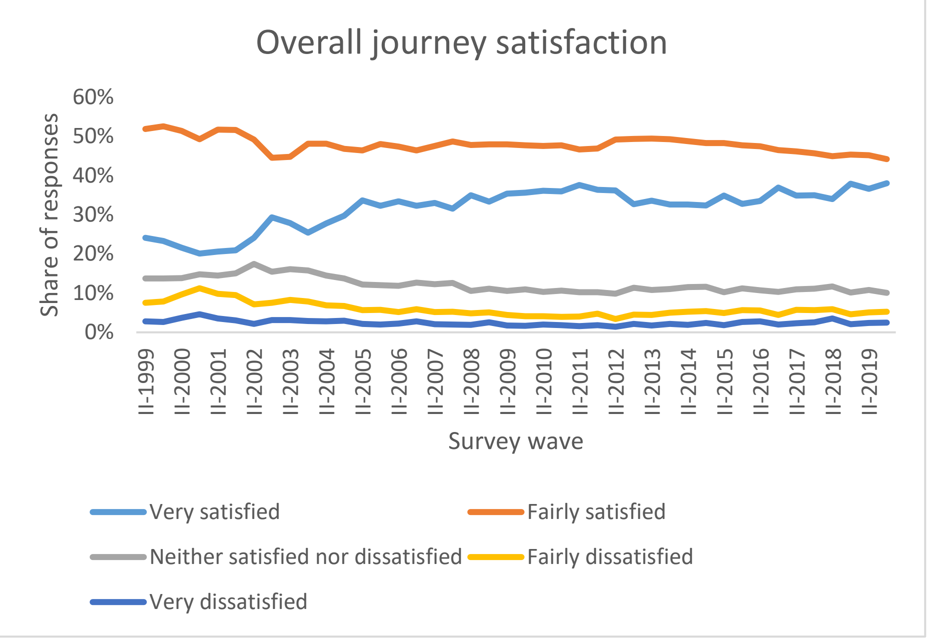
The aim of this chapter is to introduce the data used in the analysis presented in the remaining chapters of this thesis. These utilise data on satisfaction and delay perception from the National Rail Passenger Survey (NRPS) obtained from Transport Focus and matched to the operational dataset using the Historic Service Performance database obtained from Network Rail. This chapter aims to summarise the two data sources, comment on the choice of variables used throughout the thesis and describe how satisfaction and operational data were matched.

The National Rail Passenger Survey (NRPS) is conducted twice a year by Transport Focus in the United Kingdom and is concerned with rail passengers' journey satisfaction. The survey has been built into franchising agreements and provides a passenger-centric perspective on the comparative performance of franchised TOCs (Campaign for Better Transport, 2015). The results of the survey have often been cited in research papers (i.e. Oliveira et al., 2019; Ojeda-Cabral et al., 2021; Calastri et al., 2022; Smith and Ojeda Cabral, 2022) and used by researchers to study:

- passenger behaviour and use of time (Lyons et al., 2007; Lyons et al., 2016),
- differences between open-access and franchised train operators (Stead et al., 2019),
- relationship between delays and passenger satisfaction (Monsuur et al., 2021) and
- impact of train and station types on service quality perceptions (Monsuur et al., 2017).

The survey is administered by intercepting passengers in the course of making a journey and consists of multiple questions relating to passengers' satisfaction with different journey aspects (from station facilities and ticketing to journey times and in-vehicle experience). Typically, travellers receive questionnaires prior to boarding their services.

As shown in Figure 11, the overall journey satisfaction levels have generally improved since the 2000s, driven by increases in the share of the 'very satisfied' passengers. In 2019, the overall satisfaction levels by TOC varied between 72%-96% (Transport Focus, 2019). At the same time, satisfaction with punctuality had initially improved in the early 2000s, given the increase in the share of the 'very satisfied' passengers. Otherwise, the punctuality satisfaction levels have remained relatively unchanged with a similar proportion of travellers reporting to have been ‘very’ and ‘fairly satisfied’ with the remaining three satisfaction categories typically being chosen by around 5% of respondents each.



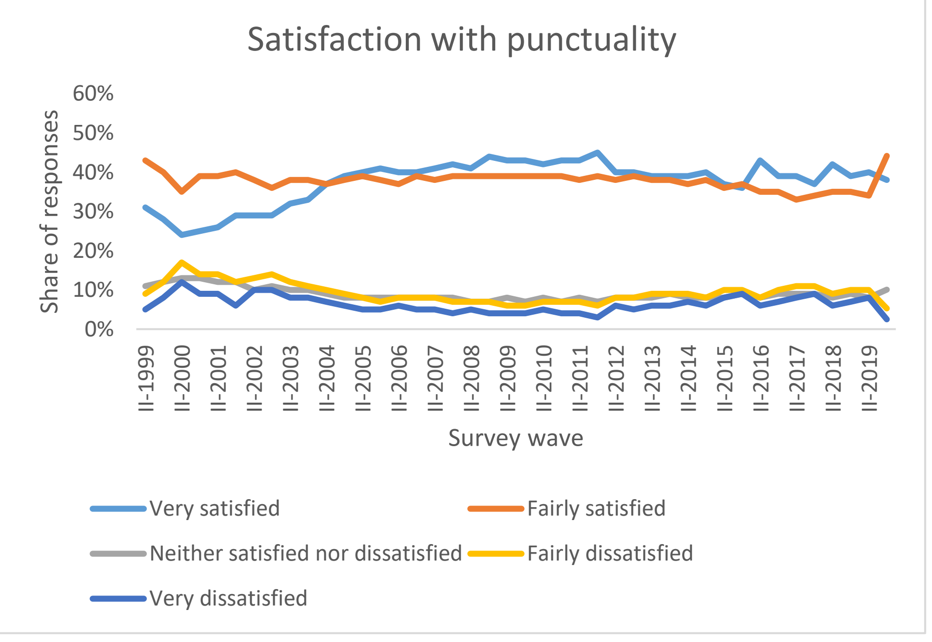


Figure 11 Evolution of passenger satisfaction levels over time based on overall journey satisfaction (top) and satisfaction with punctuality (bottom)

Public Performance Measure (PPM), a standard industrial measure of performance combining punctuality and reliability and indicating the proportion of services arriving to the destination on time, had improved since the 2000s but started worsening around 2010

(Campaign for Better Transport, 2015; ORR, 2020) as shown in Figure 12. As the figure presents performance by rail quarter, some seasonal impacts can also be visible.

As a result, Average Passenger Lateness (APL), defined as the average length of delay a typical passenger is subjected to, increased, as depicted in Figure 13 below. An average passenger journey was subjected to a delay of around 3 minutes in the most recent (pre-pandemic) years compared to around 2 minutes at the beginning of 2010s. As rail delays are, however, usually dependent on many different factors and not necessarily equally distributed across journeys, geographies, or days of the year, the NRPS survey responses are not necessarily representative of the delay distribution across the network.

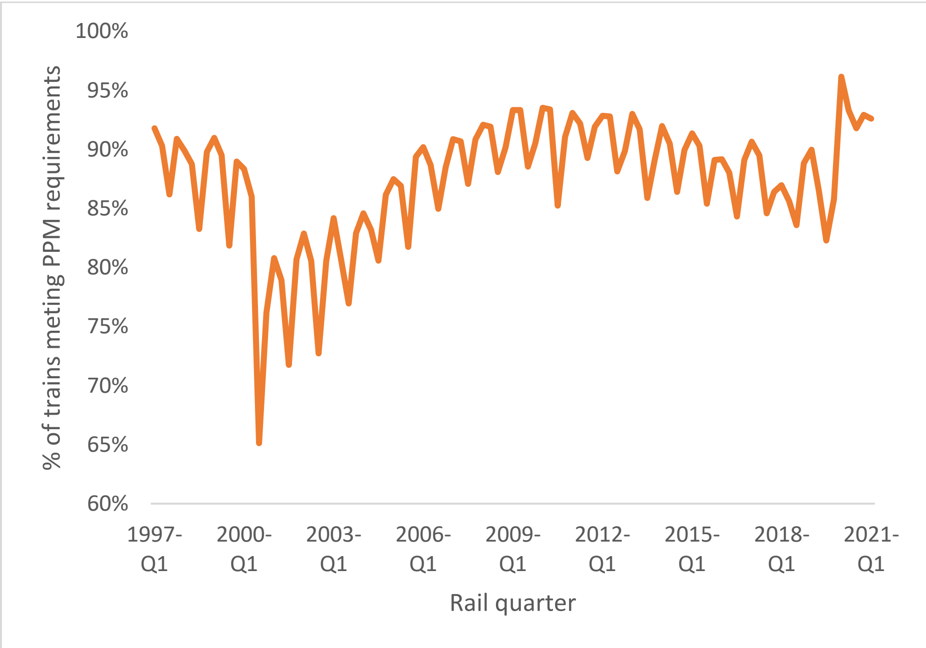


Figure 12 Rail performance measure by PPM by rail quarter (ORR, 2021)

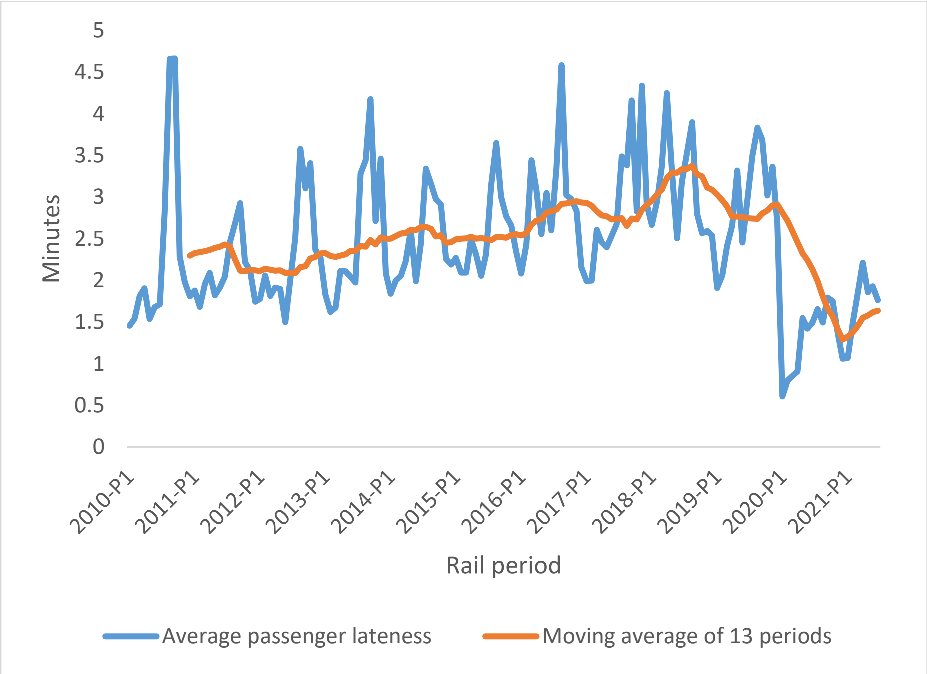


Figure 13 Average passenger lateness (APL) by rail period (ORR, 2021)

5.2. Data processing and matching passenger responses to the operational dataset

The NRPS data were obtained directly from Transport Focus and whilst it is not available open-source, anyone interested can request access by contacting Transport Focus. To some extent, a relatively detailed analysis of the data can also be conducted using the Transport Focus data hub or the reports produced by Transport Focus after each survey wave. This piece of work uses data from almost 275,000 responses from 10 survey waves (between autumn 2015 and winter 2020). Each wave of the survey typically captures around 25,000-30,000 passenger responses. The sampling design undertaken by Transport Focus ensures that different types of journeys, origin-destination pairs and passengers are represented in the sample (Transport Focus, 2020b). After filtering out the responses with missing data, the responses were subsequently matched with operational data using the Historic Service Performance (HSP) database. This platform contains historical data on train performance from Darwin (running information engine). The HSP is freely available upon registering for access using the National Rail Data Portal (for a more detailed description see National Rail Enquiries, 2021). In this way, each passenger journey was matched to an actual service and subsequently, scheduled and actual journey times were calculated for each of the journeys.

As part of the NRPS, passengers were asked to specify their departure time, origin (where the questionnaires were handed) and destination, focusing on the specific journey leg. This allowed matching each response to the actual service to calculate scheduled and actual journey times. This matching process is a two-step process (as shown in Figure 14) that involves:

1) searching for a scheduled train running between the specified origin and destination within a specified time window (i.e. the departure time stated by passenger) and
2) subsequently, a service RID code corresponding to the service that passenger travelled on is obtained and used to retrieve additional information about scheduled and actual departure and arrival times at the origin and destination stations.

Figure 14 The matching process

Elaborating on the matching process depicted in Figure 14, passenger responses were matched to the operational dataset in the following steps:

1) Origin and destination station codes were extracted from the NRPS dataset alongside the scheduled departure time and date. The data were subsequently rearranged in the format depicted in Figure 15.

	A	B	C	D	E	F	G	H
1	case_id	from_loc	to_loc	from_time	to_time	from_date	to_date	days
2	1	BIF	MIA	644	645	01/09/2015	01/09/2015	WEEKDAY
3	2	MYB	BMO	1545	1546	11/01/2016	11/01/2016	WEEKDAY
4	3	PNZ	PAD	1752	1753	02/09/2016	02/09/2016	WEEKDAY
5	4	KGX	ABD	1600	1601	01/09/2015	01/09/2015	WEEKDAY
6	5	DDG	KID	1517	1518	11/01/2016	11/01/2016	WEEKDAY
7	6	LBG	EBN	1823	1824	02/09/2016	02/09/2016	WEEKDAY
8	7	BAN	MYB	645	646	11/01/2016	11/01/2016	WEEKDAY
9	8	MAN	BRI	944	945	01/09/2015	01/09/2015	WEEKDAY
10	9	BHM	LIV	1636	1637	11/01/2016	11/01/2016	WEEKDAY
11	10	DID	PAD	749	750	02/09/2016	02/09/2016	WEEKDAY

Figure 15 Dataset snapshot

2) The data described under point 1 were then used to retrieve a service RID code corresponding to the service that a passenger travelled on between the specified origin and destination as shown in Figure 16-Figure 18.

```
In [6]: print(A, B, C, D, E, F, G)
KGX ABD 1600 1601 2015-09-01 2015-09-01 WEEKDAY
```

Figure 16 NRPS data matching input

```
In [5]: print(resp.content)
b'{"header":{"from_location":"KGX","to_location":"ABD"},"Services":[{"serviceAttributesMetrics":
{"origin_location":"KGX","destination_location":"ABD","gbtt_ptd":"1600","gbtt_pta":"2312","toc_code":
"GR","matched_services":"1","rids":["201509010384368"]},"Metrics":
[{"tolerance_value":"0","num_not_tolerance":"0","num_tolerance":"1","percent_tolerance":"100","global
_tolerance":true}]}]}'
```

Figure 17 HSP output

```
In [7]: print(rid)
201509010384368
```

Figure 18 Matched RID code

3) The RID codes obtained were then used to request service details information as shown in Figure 19. In this example, a London Kings Cross to Aberdeen service with an RID code of 201509010384368 was scheduled to depart from London Kings Cross at 16:00 on the 1st of September 2015. The service was scheduled to arrive at Aberdeen at 23:12 after having stopped at 13 intermediate stations. It departed from London Kings Cross a minute earlier than scheduled (15:59) and arrived at the destination 4 minutes ahead of schedule (23:08).

```
In [9]: print(resp.content)
```


```
b'{"serviceAttributesDetails":
{"date_of_service":"2015-09-01","toc_code":"GR","rid":"201509010384368","locations":
[{"location":"KGX","gbtt_ptd":"1600","gbtt_pta":"","actual_td":"1559","actual_ta":"","late_canc_reaso
n":""},
{"location":"YRK","gbtt_ptd":"1754","gbtt_pta":"1752","actual_td":"1756","actual_ta":"1753","late_can
c_reason":""},
{"location":"DAR","gbtt_ptd":"1823","gbtt_pta":"1821","actual_td":"1824","actual_ta":"1822","late_can
c_reason":""},
{"location":"NCL","gbtt_ptd":"1855","gbtt_pta":"1851","actual_td":"1855","actual_ta":"1849","late_can
c_reason":""},
{"location":"BWK","gbtt_ptd":"1941","gbtt_pta":"1939","actual_td":"1942","actual_ta":"1939","late_can
c_reason":""},
{"location":"EDB","gbtt_ptd":"2032","gbtt_pta":"2022","actual_td":"2032","actual_ta":"2023","late_can
c_reason":""},
{"location":"HYM","gbtt_ptd":"2037","gbtt_pta":"2035","actual_td":"2037","actual_ta":"2035","late_can
c_reason":""},
{"location":"INK","gbtt_ptd":"2053","gbtt_pta":"2051","actual_td":"","actual_ta":"","late_canc_reason
":""},
{"location":"KDY","gbtt_ptd":"2111","gbtt_pta":"2109","actual_td":"","actual_ta":"","late_canc_reason
":""},
{"location":"LEU","gbtt_ptd":"2139","gbtt_pta":"2137","actual_td":"2139","actual_ta":"2135","late_can
c_reason":""},
{"location":"DEE","gbtt_ptd":"2153","gbtt_pta":"2152","actual_td":"2153","actual_ta":"2151","late_can
c_reason":""},
{"location":"ARB","gbtt_ptd":"2211","gbtt_pta":"2209","actual_td":"2212","actual_ta":"2210","late_can
c_reason":""},
{"location":"MTS","gbtt_ptd":"2227","gbtt_pta":"2225","actual_td":"2228","actual_ta":"2225","late_can
c_reason":""},
{"location":"STN","gbtt_ptd":"2250","gbtt_pta":"2248","actual_td":"","actual_ta":"","late_canc_reason
":""},
{"location":"ABD","gbtt_ptd":"","gbtt_pta":"2312","actual_td":"","actual_ta":"2308","late_canc_reason
":""}]}}'
```


Figure 19 HSP service details

4) The origin and destination station codes were subsequently used to retrieve scheduled and actual departure and arrival times for the specified origin and destination pair.

5) Finally, scheduled (JT_S) and actual journey lengths (JT_A), and delay lengths at departure (L_D) and arrival (L_A) were calculated. For clarity, the definitions are shown below:

Scheduled journey length (JT_S) is the difference between the scheduled arrival time at the destination station ($Arrival_S$) and scheduled departure time at the origin station ($Departure_S$).

$$JT_S = Arrival_{\,S} - Departure_S \tag{5}$$

Delay length at arrival (L_A) is the difference between the actual arrival time at the destination station ($Arrival_A$) registered in the Historic Service Performance database and scheduled arrival time at the destination station ($Arrival_S$).

$$L_A = Arrival_{\,A} - Arrival_{\,S} \tag{6}$$

Delay length at departure (L_D) is the difference between the actual departure time at the origin station ($Departure_A$) registered in the Historic Service Performance database and scheduled departure time at the origin station ($Departure_S$).

$$L_D = Departure\ _A - Departure\ _S \tag{7}$$

5.2.1. Summary of the matching process and issues encountered

Out of the 274,862 responses in the dataset, 263,163 (95.7%) responses with no missing data were selected for matching with operational data. 244,712 (93.0%) were matched to an actual service RID code by searching for a service that was scheduled to depart from the origin station at the departure time stated by passenger.

Transport Focus was approached to discuss the possible reasons for some of the responses not being matched with the operational data. As noted, it is expected that in some cases passengers rounded the departure times or provided an erroneous departure time based on actual (or expected) rather than scheduled departure time. Considering the proportion of responses successfully matched to the operational dataset, 93% was suggested to be in line with what was achieved in similar attempts conducted in the past.

Following these considerations, an attempt was made to find possible services that could match the responses where a service RID code was not matched automatically. It was, however, found that while in some cases the differences between passenger-stated and scheduled departure times of a service are relatively small, there was still a considerable number of responses where finding a scheduled service based on passenger-stated scheduled departure time was more challenging. Therefore, it was decided to extend the search window by respectively +/- 3, 5 and 10 minutes from the stated departure time. The matching algorithm is depicted in Figure 20.

By doing that an additional 1.4%, 1.9% and 2.5% responses were matched to an actual service. It was thought that extending the search window by +/- 5 minutes was possibly most valid as it minimises the risk of (erroneously) matching a different service, which can be the case if the search window is extended by +/- 10 minutes as on the busier stations headway can often be less than 20 minutes.

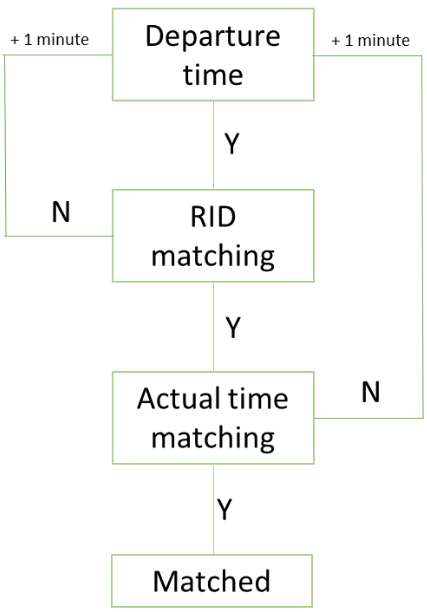


Figure 20 The matching algorithm

Through this approach, the final dataset consisted of 249,686 responses successfully matched to a service RID code. Subsequently, service details were successfully retrieved for 242,311 responses as summarised in Table 14.

Table 14 Summary of the HSP matching process

	Responses matched	Total responses matched (%)
All responses with a satisfaction score	263,163	100.00%
	Matching service RID codes	
RID matched	244,712	92.99%
	Extending search window	
+/- 3 minutes	3,685	+1.40%
+/- 5 minutes	4,974	+1.89%
+/- 10 minutes	6,450	+2.45%
	Final dataset	
Passenger-stated departure time +/- 5 minutes	249,686	94.88%
	Matching actual running times	
Actual running times matched	242,311	92.08%

When an RID code was matched, but no actual departure and arrival times were found, this may have been due to recording errors. It was initially thought that this might be due to cancellations or truncation of the services. An investigation was conducted to better understand how cancellations or truncations are registered in the database and find the possible reasons for some service code RIDs being found, but not matched to the operational data. At first, a user forum, which is a place for an unofficial exchange of information and community-based support with occasional support provided by the Rail Delivery Group, was consulted to see if other users previously raised similar issues. One important consideration raised by the Rail Delivery Group representative was that HSP only reports on locations that have an actual running time or were cancelled. In the cases where no movement report was received and the cancellation state was not manually applied, the service becomes invisible for the HSP. Therefore, the two possible reasons for service details not being matched to the RID code supplied are errors with recording the train movement data or errors with manual inputs of cancellations.

To better understand the mechanisms of HSP, an example of a disruption to the services was chosen for illustration. On the 13th of September 2021 some services were cancelled to/from Ilkley (ILK) following an emergency incident (Ilkley Gazette, 2021). Out of the 11 services investigated, five were cancelled. Figure 21 and Figure 22 show how cancellation was recorded in the HSP database for the 06:51 Ilkley to Bradford Foster Square (BDQ) service and 09:03 Leeds (LDS) to Ilkley service respectively. The first service departed from Ilkley a minute ahead of schedule and the last station served was Guiseley where the service arrived a minute ahead of schedule. For the remaining four stations, planned departure and arrivals are supplied, but the actual recorded times are missing. In this case, a late/cancellation reason code (777) was also supplied. In the case of the LDS-ILK 09:03 service, it was already cancelled at Leeds and did not depart from the origin station. Similarly, a late/cancellation reason code (777) was also supplied in this case.

```
b'{"serviceAttributesDetails":
{"date_of_service":"2021-09-13","toc_code":"NT","rid":"202109137607600","locations"
:
[{"location":"ILK","gbtt_ptd":"0651","gbtt_pta":"","actual_td":"0650","actual_ta":"
","late_canc_reason":"777"},
{"location":"BEY","gbtt_ptd":"0653","gbtt_pta":"0653","actual_td":"0654","actual_ta
":"0652","late_canc_reason":"777"},
{"location":"BUW","gbtt_ptd":"0658","gbtt_pta":"0658","actual_td":"0659","actual_ta
":"0657","late_canc_reason":"777"},
{"location":"MNN","gbtt_ptd":"","gbtt_pta":"0701","actual_td":"","actual_ta":"0659"
,"late_canc_reason":"777"},
{"location":"GSY","gbtt_ptd":"0705","gbtt_pta":"0705","actual_td":"","actual_ta":"0
704","late_canc_reason":"777"},
{"location":"BLD","gbtt_ptd":"0711","gbtt_pta":"0710","actual_td":"","actual_ta":""
,"late_canc_reason":"777"},
{"location":"SHY","gbtt_ptd":"0714","gbtt_pta":"0714","actual_td":"","actual_ta":""
,"late_canc_reason":"777"},
{"location":"FZH","gbtt_ptd":"0717","gbtt_pta":"0717","actual_td":"","actual_ta":""
,"late_canc_reason":"777"},
{"location":"BDQ","gbtt_ptd":"","gbtt_pta":"0723","actual_td":"","actual_ta":"","la
te_canc_reason":"777"}]}}'
```


Figure 21 HSP service details outputs for the ILK-BDQ service


```
b'{"serviceAttributesDetails":
{"date_of_service":"2021-09-13","toc_code":"NT","rid":"202109137609952","locations"
:
[{"location":"LDS","gbtt_ptd":"0903","gbtt_pta":"","actual_td":"","actual_ta":"","l
ate_canc_reason":"777"},
{"location":"GSY","gbtt_ptd":"","gbtt_pta":"0914","actual_td":"","actual_ta":"","la
te_canc_reason":"777"},
{"location":"MNN","gbtt_ptd":"0918","gbtt_pta":"0918","actual_td":"","actual_ta":""
,"late_canc_reason":"777"},
{"location":"BUW","gbtt_ptd":"0921","gbtt_pta":"0920","actual_td":"","actual_ta":""
,"late_canc_reason":"777"},
{"location":"BEY","gbtt_ptd":"0924","gbtt_pta":"0924","actual_td":"","actual_ta":""
,"late_canc_reason":"777"},
{"location":"ILK","gbtt_ptd":"","gbtt_pta":"0931","actual_td":"","actual_ta":"","la
te_canc_reason":"777"}]}}'
```


Figure 22 HSP service details outputs for the LDS-ILK service

As shown in the matching algorithm, an attempt was also made to search for the closest service that had the actual running data available for the services where the RID code was found, but the service details were not successfully retrieved. If a new RID code was found, the algorithm proceeded to match actual running times. If the actual running times were not found, the algorithm returned to finding the next possible service by increasing departure time by 1 minute until a new RID and actual running times were subsequently matched. The cancellation algorithm was capped at 6 hours. In some cases, specific OD pairs with no services matching passenger journey were investigated manually. In the case of journeys between Shanklin (SHN) and Ryde Pier Head (RYP) (depicted in Figure 23), no services were found in the HSP on the 19th of February 2020, but the service details were successfully retrieved for the week earlier. This would indicate on some possible errors with the recording system, but the magnitude of that issue is relatively small.

c4_actu	date	daytype	daytype	origin	destinat
1018	27/02/2020	Weekday	Weekday	SHN	RYP
1038	02/03/2020	Weekday	Weekday	SHN	RYP
907	07/10/2019	Weekday	Weekday	RYP	SHN
1038	07/10/2019	Weekday	Weekday	SHN	RYP
618	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
918	19/02/2020	Weekday	Weekday	SHN	RYP
718	19/02/2020	Weekday	Weekday	SHN	RYP

Figure 23 SHN-RYP services

In conclusion, filtering out the responses where some data were missing and the responses that were not successfully matched to the operational data, 242,311 responses (92% of the original dataset) were selected for further investigation.

5.2.2. Recorded and stated delay lengths

While the delay data were obtained by matching the individual responses to the operational data using the HSP database as described in the previous section, the survey also contained a question regarding late running. This is used for monitoring the quality of the data by comparing the delay lengths recorded in the HSP database with the delays reported by passengers. Moreover, the comparison of the two sources can help better understand passengers' ability to perceive delays (this will be explored as part of Chapter 6).

In the first three waves of the survey, passengers were asked: "*How long was your delay?*" and needed to state their delay lengths in minutes. This was later changed and in the most recent seven versions of the survey, passengers needed to choose the length category their delay length fell into as shown in Figure 24. The survey data on delay lengths was used for monitoring and comparison purposes with the delay categories converted into delay minutes using the midpoint method and assuming the average delay of 90 minutes for the over 60 minutes delay length category.

Q13 **Did you experience any delay either on this train or because the train you had planned to catch at Glasgow Central was cancelled?**

No delay.......... ☐ **Go to Q16**
Up to 5 minutes delay.......... ☐ **Go to Q14**
6-10 minutes delay.......... ☐ **Go to Q14**
11-15 minutes delay.......... ☐ **Go to Q14**
16-20 minutes delay.......... ☐ **Go to Q14**
21-30 minutes delay.......... ☐ **Go to Q14**
31-60 minutes delay.......... ☐ **Go to Q14**
Over 60 minutes delay.......... ☐ **Go to Q14**

Figure 24 Question about the perception of delay from NRPS (Transport Focus, 2020)

As shown in Table 15 below, almost 1 in 2 surveyed passengers arrived at their destination late according to the operational data, but only 1 in 5 actually reported being delayed. The matching methodology is expected to be more accurate than passenger perception data. In this case, a recorded delay is defined as a difference between actual and scheduled arrival time at the destination. This way, the smallest possible delay recorded is 1 minute. It is, however, unlikely that passengers are able to perceive the smaller delays and this is the potential explanation for a larger number of responses being matched to a delay than the number of passengers reporting late arrival. Nevertheless, it is also important to recognize smaller (likely unperceived) delays and this aspect is further explored as part of this thesis. In 6% of responses, a passenger claimed they were delayed, but no delay was matched. It is possible that in these cases passengers were either not able to board the train due to crowding or included the delay from a different leg of the journey or due to missed connections. In this case, it is not possible to trace the whole journey that a passenger was intending to make and these responses are discarded at this stage.

Table 15 Comparison between passenger-stated and HSP-matched delay data

		Perceived	
		Delay	**No delay**
Recorded	**Delay**	18.0%	30.8%
	No delay	5.8%	45.3%

Typically, respondents failed to report shorter delays - possibly due to shorter delays falling below their perceptual thresholds. An average passenger lateness of 2.7 minutes was recorded for the whole NRPS sample. For the subset of journeys where a delay was recorded, an average delay of 3.6 minutes was recorded for passengers who were matched to a delay but did not report late running (i.e. row 1, column 2 in Table 15) versus 9.0 minutes for passengers who also reported being delayed (i.e. row 1, column 1 in table 2). Figure 25 shows the cumulative distribution of recorded delays for passengers who did and did not report late running. Out of the 30.8% (74,628) responses, in almost 70% of cases, the matched delay length was within 3 minutes and in 98% of cases within 15 minutes. In line with expectations, it can be seen that passengers who were matched to a delay, but reported arriving on time (i.e. row 1, column 2 in Table 15), were typically matched a shorter delay.

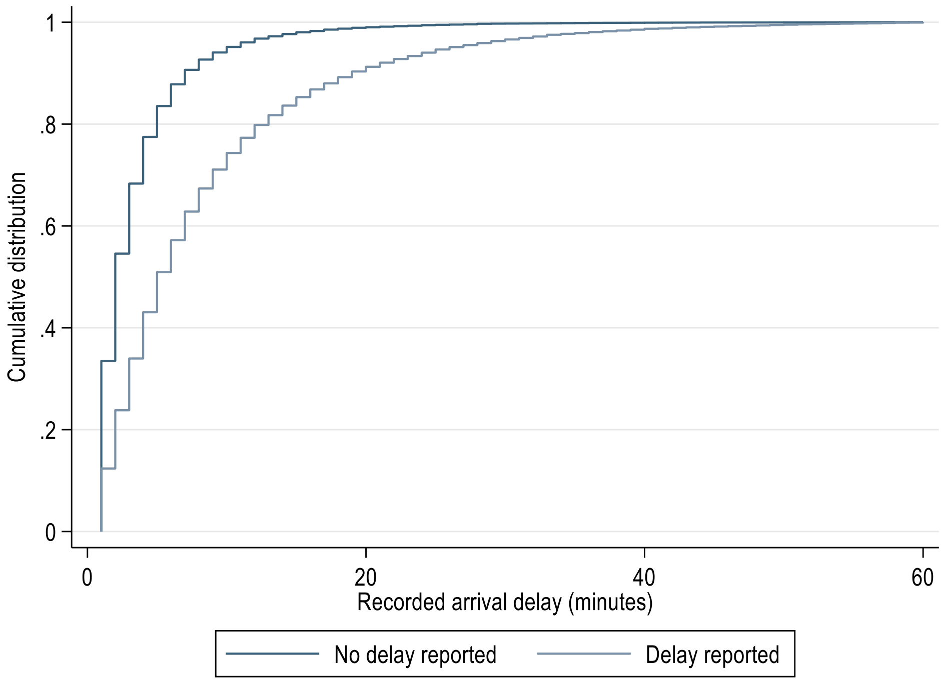


Figure 25 Cumulative distribution of recorded delays for perceived and unperceived delays

At the same time, Figure 26 shows how the proportion of passengers reporting late running changes with increasing delay lengths. It is worth noting that the number (not proportion) of responses decreases with increasing delay lengths, as typically most delays are relatively small and the delay distribution is positively skewed. The proportion of passengers reporting late running seems to be generally increasing with delay length, but this relationship needs to be studied in more detail. As suggested by Monsuur et al. (2021), 30 minutes is a delay length following which passengers are very unlikely to remain satisfied with their journey. Therefore, the dataset is constrained to recorded delay lengths of up to 30 minutes as it is assumed that passengers should already be able to perceive delays of that length and any discrepancies (i.e. passengers not reporting delays despite a recorded delay of over 30 minutes) may be erroneous (as described previously). Moreover, delays of over 30 minutes are relatively rare and the main focus of this study is on the smaller delays.

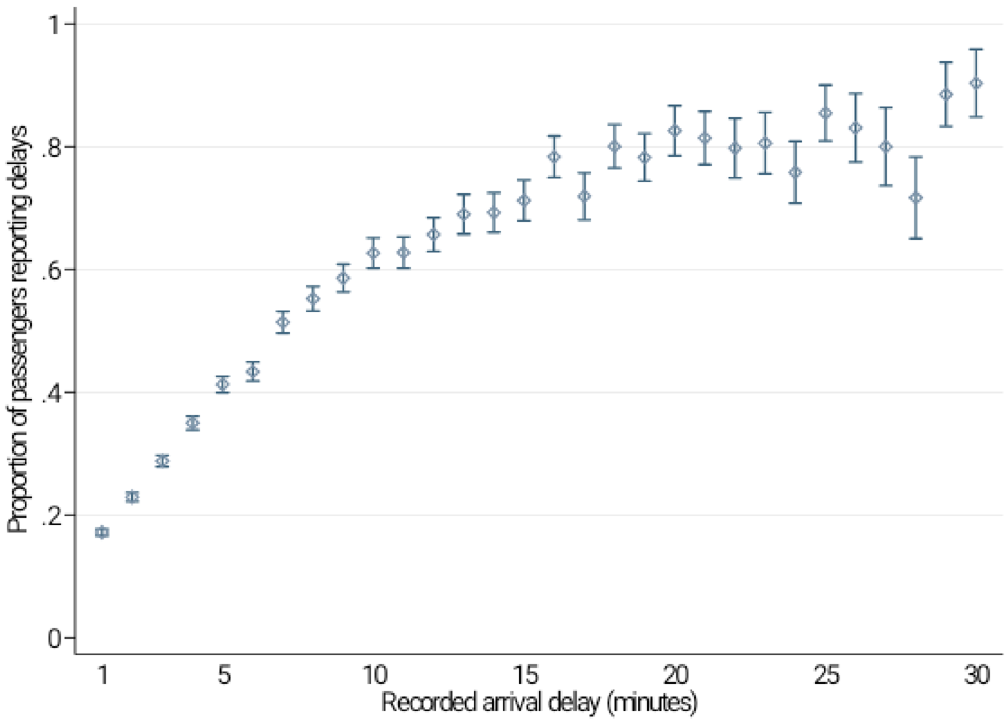


Figure 26 The proportion of passengers reporting late running

As shown in Table 15, almost 2 out of 3 passengers agreed with the operational data regarding whether or not they were delayed. The cases where passengers failed to perceive the smaller recorded delays or reported delays when on-time performance was recorded are discussed above. However, it remains to be investigated how the reported delay lengths compare to the matched delay lengths for the cases where passengers who were matched to a delay also reported late arrival (i.e. row 1, column 1 in Table 15). Some discrepancies are expected and may be a result of:

- passengers rounding the delay lengths to the closest 5 or 10 minutes (in the case of the first three survey waves where they were asked to type in the exact delay length) or
- the conversion method for the delay length categories used in the latter 7 survey waves as described before.

An attempt was made to understand the extent of these differences with the summary presented in Table 16 below. However, a more detailed analysis will be conducted as part of the investigation of delay perception in Chapter 6. In 11% of cases, there was no difference between stated and matched delay lengths with over 2/3 of the differences being within 5 minutes and only a small percentage (i.e. around 9%) of responses being characterised by the differences in matched and reported delay lengths of over 15 minutes.

Hence, it can be expected that most of the differences can be attributed to passenger perception.

Table 16 Summary of differences between stated and matched delay lengths

Difference	Percentage of delayed responses
No difference	10.6%*
Within +/- 3 minutes	57.2%*
Within +/- 5 minutes	70.1%*
Within +/- 15 minutes	91.2%*
Within +/- 30 minutes	96.3%*
>30 minutes	3.7%

*Cumulative distribution

Given the suggestions that departure L_D and arrival L_A delays have a differing impact on passengers (see Batley and Ibáñez, 2012 for a more detailed discussion), this research also aims to investigate the impact that both can have on delay perception and satisfaction. Departure (origin) and arrival (destination) delays are, as expected, correlated (r=0.64) in the sample. Whilst in most cases throughout the thesis, the focus remains on the arrival delay, in some cases both variables are used in the modelling.

5.3. The National Rail Passenger Survey

The obtained data comes from 10 survey waves of NRPS between Autumn 2015 and Winter 2020. The sample design and the weighting process conducted by Transport Focus ensure that the responses are distributed across the different operators and routes over the different times of day and days of the week (Transport Focus, 2020). The sampling process is described in technical reports produced for each of the survey waves and generally involves the following steps:

1) The whole network is divided into multiple building blocks.
2) Selection of stations for each of the building blocks uses a PPS (probability proportionate to size) basis, so that the sample sizes are adjusted for station usage.
3) Day of the week and times of the day distribution is based on the profiles of journey departures by journey purpose provided by TOCs.
4) Sampling points are then assigned to weeks at random during the survey period.

The questionnaires consist of multiple questions relating to the information about a passenger, the specific journey that the passenger made and their satisfaction with different aspects of that journey. The following subsections aim to review the different types of questions asked as part of the survey that are particularly relevant to the analysis conducted as part of the thesis.

5.3.1. Overall journey satisfaction and satisfaction with punctuality

This work primarily focuses on analysing the impacts of increasing delays on delay perception and satisfaction. Therefore, there are two questions that appear particularly relevant in this context:

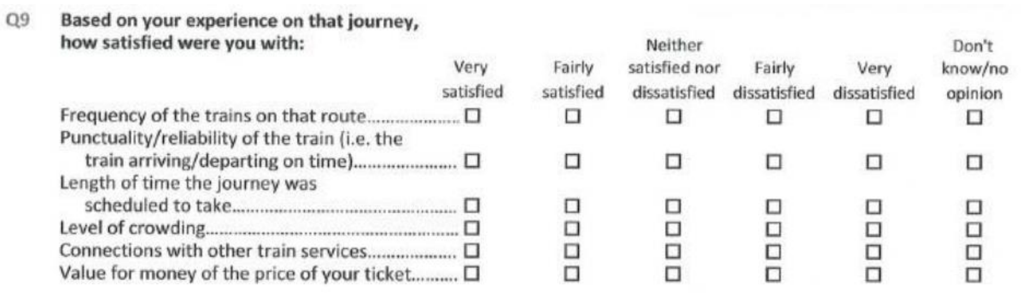
1) Question 16 of the NRPS that concerns passengers' overall satisfaction with their journey (Figure 27 below). Passengers are specifically asked to score their satisfaction with the origin station and the train they boarded after receiving a questionnaire.

4 Your overall opinion of your journey today

	Very satisfied	Fairly satisfied	Neither satisfied nor dissatisfied	Fairly dissatisfied	Very dissatisfied	Don't know/no opinion
Q16 Taking into account Glasgow Central station where you boarded the train and the actual train travelled on after being given this questionnaire, how satisfied were you with your journey today?	☐	☐	☐	☐	☐	☐

Figure 27 Overall satisfaction question from NRPS

2) An alternative line of enquiry is to consider NRPS question 9, which is concerned specifically with satisfaction with punctuality and reliability of the train used (Figure 28). It potentially allows for direct analysis of the delay impacts without the need to control for satisfaction with other aspects of journey quality.



Q9 Based on your experience on that journey, how satisfied were you with:

	Very satisfied	Fairly satisfied	Neither satisfied nor dissatisfied	Fairly dissatisfied	Very dissatisfied	Don't know/no opinion
Frequency of the trains on that route....................	☐	☐	☐	☐	☐	☐
Punctuality/reliability of the train (i.e. the train arriving/departing on time)......................	☐	☐	☐	☐	☐	☐
Length of time the journey was scheduled to take..	☐	☐	☐	☐	☐	☐
Level of crowding...	☐	☐	☐	☐	☐	☐
Connections with other train services....................	☐	☐	☐	☐	☐	☐
Value for money of the price of your ticket..........	☐	☐	☐	☐	☐	☐

Figure 28 Question related to satisfaction with punctuality from NRPS

Passengers scored their overall satisfaction as well as satisfaction with punctuality on a 5-point Likert scale with possible responses ranging from 'very satisfied' (5) to 'very dissatisfied' (1) and a 'don't know/no opinion' option. Out of the 242,311 responses chosen for further analysis, 6,596 passengers (2.7%) chose the 'don't know/no opinion' option regarding their satisfaction with punctuality (such responses have been excluded from the analysis). Monsuur et al. (2021) used overall journey satisfaction (i.e. question

16) to study the impacts of delays on passenger satisfaction. To better understand the difference between overall journey satisfaction and satisfaction with punctuality and to determine their usefulness, the two variables were investigated more closely.

It is expected that delays have a negative impact on journey satisfaction. While the two satisfaction variables are correlated (i.e. Spearman rho of 0.57), it is also expected that satisfaction with punctuality is only one of the many factors determining overall journey satisfaction. Table 17a below shows how average recorded delay and overall satisfaction (NRPS question 16 in Figure 27) change with reported satisfaction with punctuality (NRPS question 9 in Figure 28) for travellers with a recorded delay. As expected, passengers who scored their satisfaction with punctuality lower were typically subjected to lengthier delays – from around 4 minutes of average recorded delay for passengers 'very satisfied' with punctuality to 12 minutes for those who were 'very dissatisfied'. The overall satisfaction levels (i.e. NRPS Q16) decrease with both increasing delays and decreasing satisfaction with punctuality (i.e. NRPS Q9) – from 4.6 for passengers 'very satisfied' with punctuality to 2.5 for passengers 'very dissatisfied' with punctuality. Table 17b, in turn, shows how the same relationship changes for decreasing levels of reported overall journey satisfaction.

Table 17a Relationship between satisfaction with punctuality, overall journey satisfaction and delay lengths (for journeys with a matched delay only)

Satisfaction with punctuality	Average recorded delay	Number of responses	Average overall satisfaction
Very satisfied	4.07	42,632	4.58
Fairly satisfied	4.64	38,989	4.06
Neither satisfied nor dissatisfied	5.92	10,052	3.71
Fairly dissatisfied	8.23	13,806	3.38
Very dissatisfied	11.99	9,754	2.50
Total	5.59	115,233	4.01

Table 17b Relationship between overall journey satisfaction, satisfaction with punctuality and delay lengths (for journeys with a matched delay only)

Overall journey satisfaction	Average recorded delay	Number of responses	Average punctuality satisfaction
Very satisfied	4.33	39,954	4.61
Fairly satisfied	5.24	54,460	3.78
Neither satisfied nor dissatisfied	6.61	12,909	2.82
Fairly dissatisfied	9.91	7,542	2.15
Very dissatisfied	12.61	3,438	1.72
Total	5.59	115,233	3.79

5.3.2. Control variables

Sixty three different questions related to satisfaction with very specific journey aspects along three more general questions were asked throughout the 10 NRPS survey waves used in this study. Satisfaction with each aspect of the journey was scored by passengers on a 5-point Likert scale with an option of no score if a passenger felt they did not know how to score this aspect or did not experience/use it (e.g. toilet facilities, catering, etc.). The possibility to choose the option where no score was provided means that a large proportion of questions is characterised by a rather small percentage of responses. Moreover, some questions are mutually exclusive (e.g. a person who uses a car park is unlikely to also use a bike park). While many of these questions had a relatively low response (i.e. large proportion of passengers choosing an option not to score a specific element) or were not part of all the survey waves, 16 were chosen for further investigation based on the common number of responses with a summary presented in Table 18 below.

Out of the 242,311 responses where a passenger scored their overall journey experience, there were 126,794 responses where a score was provided for all of the 16 questions relating to the specific aspects of the journey and three additional questions related to satisfaction with train, station, and overall journey as detailed in Table 19. Similarly as observed by Brons and Rietveld (2009), the average overall satisfaction is in most cases higher than the average satisfaction with the specific journey aspects.

In cases where the overall journey satisfaction is modelled, there is a need to control for other aspects of journey satisfaction too (i.e. overall satisfaction does not only depend on the length of delay). In particular, it can be expected that the overall journey satisfaction is impacted by satisfaction with journey quality aspects (e.g. station or train), satisfaction with journey frequency, punctuality, scheduled journey time and/or value for money. However, in the case where satisfaction with punctuality is modelled, the quality aspects are likely to have a complementary (to delay length), but not a direct impact on punctuality satisfaction as bad journey quality (e.g. crowding) may amplify the negative impact of delay on satisfaction. Table 20 below reports correlations between the key satisfaction variables using Spearman's rank correlation methodology (Spearman, 1904).

Table 18 Choice of satisfaction variables from NRPS

Aspect	Question
Train and platform information	How would you rate this station for provision of information about train times/platforms?
Station upkeep	How would you rate this station for the upkeep/repair of the station buildings/platforms?
Station cleanliness	How would you rate this station for cleanliness of the station?
Station environment	How would you rate this station for the overall station environment?
Station security	How would you rate this station for your personal security whilst using this station?
Station seating	How would you rate this station for availability of seating?
Journey frequency	Based on your experience on that journey, how satisfied were you with the frequency of the trains on that route?
Delay	Based on your experience on that journey, how satisfied were you with the punctuality/reliability (i.e. the train arriving/departing on time)?
Scheduled journey time	Based on your experience on that journey, how satisfied were you with the length of time the journey was scheduled to take?
Value for money	Based on your experience on that journey, how satisfied were you with the value for money of the price of your ticket?
Train security	How would you rate the train you boarded for that journey in terms of your personal security whilst on board the train?
Train upkeep	How would you rate the train you boarded for that journey in terms of upkeep and repair of the train (condition of seats, walls, tables, etc.)?
Train information	How would you rate the train you boarded for that journey in terms of the provision of information during the journey?
Train cleanliness	Specifically thinking about the cleanliness of the train you boarded for that journey, how would you rate it for the cleanliness of the inside of the train?
Train cleanliness out	Specifically thinking about the cleanliness of the train you boarded for that journey, how would you rate it for the cleanliness of the outside of the train?
Train seating comfort	How would you rate the train you boarded for that journey in terms of the comfort of the seating area?
Station overall	Overall how satisfied are you with this station?
Train overall	Overall how satisfied are you with the train you boarded for your journey?
Overall	Taking into account just the station where you boarded the train and the actual train travelled on after being given this questionnaire, how satisfied were you with your journey today?

Table 19 Summary of satisfaction variables

Dimension	Journey aspect	Responses	Mean	SD
Station quality	Train and platform information	231,419	4.22	0.872
	Station upkeep	231,830	3.95	0.925
	Station cleanliness	232,500	4.03	0.881
	Station environment	233,162	3.95	0.873
	Station security	213,454	4.03	0.831
	Station seating	217,089	3.36	1.215
Timetable	Journey frequency	233,963	3.98	1.060
	Punctuality*	235,715	4.03	1.164
	Scheduled journey time	234,299	4.24	0.923
	Value for money	222,716	3.27	1.313
Train quality	Train security	221,050	4.08	0.833
	Train upkeep	235,233	3.95	0.953
	Train information	220,333	3.98	0.970
	Train cleanliness	239,497	3.99	0.928
	Train cleanliness out	209,085	3.94	0.901
	Train seating comfort	231,137	3.78	1.009
Station overall		237,315	4.12	0.833
Train overall		240,525	4.02	0.906
Overall*		242,311	4.14	0.906

* indicates the variables used as dependent variables (described in section 5.3.1)

Table 20 Correlation between the key satisfaction variables

	(1) Overall	(2) Station	(3) Train	(4) Punctuality
(1) Overall	1.00			
(2) Station overall	0.53	1.00		
(3) Train overall	0.74	0.50	1.00	
(4) Punctuality	0.57	0.38	0.50	1.00

The satisfaction variables are expected to be correlated. In line with expectations, satisfaction with train and station is more positively correlated with the overall journey satisfaction than satisfaction with punctuality. One of the features of delays in railways is that they are typically correlated with increased crowding that, in turn, may have a negative impact on the satisfaction with train and/or station. Similarly, it is possible that travellers facing delays report lower satisfaction with other journey aspects to express general discontent related to a delay experience. However, it remains difficult to understand and isolate the effects of the general discontent from the actual experience.

The correlations between all the satisfaction variables presented in Table 19 were investigated using Spearman's rank correlation coefficients and are shown in the heat map in Figure 29. It is evident that some variables (i.e. station upkeep, station environment or cleanliness) are conceptually similar to each other, resulting in a relatively high correlation. At the same time, the overall journey satisfaction is typically more correlated to the variables relating to satisfaction with train or journey (timetable) rather than station.

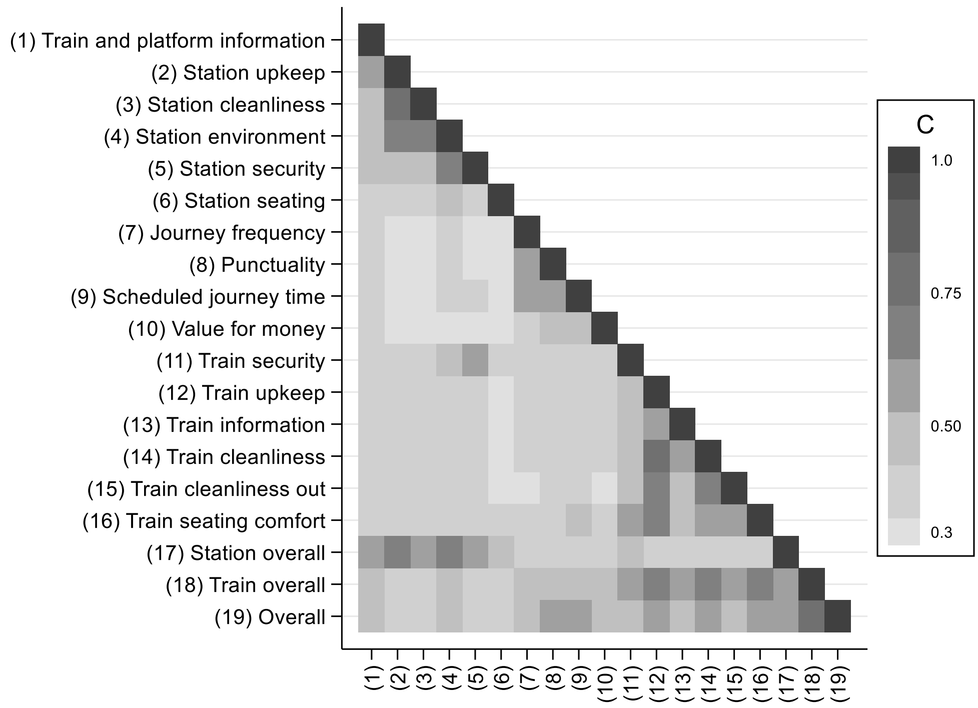


Figure 29 Heat map of correlation between key satisfaction variables

5.3.3. Other control variables

Apart from scoring satisfaction with train or seating comfort, passengers in the most recent seven survey waves were asked if they were able to find a seat on a train. 10.5% of the respondents in the sample reported that they were unable to find a vacant seat for the whole journey. It can be expected that the availability of seating has a major impact on satisfaction with the journey and may also possibly have a complementary effect on satisfaction with punctuality. An average punctuality satisfaction of 3.3 was recorded for passengers without a seat while an APL of 3.1 was registered for them. Passengers who were able to find a seat scored their punctuality satisfaction on average better (mean satisfaction of 4.1 for an APL of 2.7 minutes).

5.4. Demand segmentation

The NRPS sample design ensures that the responses of passengers travelling on all the TOCs are captured, thus the dataset contains responses from different types of travellers on both shorter and longer journeys. The average scheduled journey time in the sample is 53 minutes with an average passenger lateness of 2.2 minutes. The average passenger lateness in the sample is consistent with the values estimated by ORR (2020), suggesting APLs of 2-3 minutes in the last 10 years.

The analysis employs passenger segmentation to enable comparison of delay perception and its impacts on the different types of passengers. The proposed approach to segmentation is based on ticket types, journey purposes and geographies. An alternative to this approach could be sought in clustering passenger types based on the available data. However, the motivation behind such segmentation is to align more with the PDFH framework. Nevertheless, some simplifications to the demand segmentation typically used within the PDFH are made to facilitate the analysis (i.e. reducing the number of passenger types) whilst also replacing the journey length categorisation by the inclusion of a continuous variable representing scheduled journey time in the conducted analysis. Twelve different journey purpose categories are used in the NRPS questionnaire, similar to that in the National Travel Survey. These are then classified by the three major journey purpose categories, i.e. commute (41%), business (14%) or leisure (45%), with the split being similar to that suggested in the National Travel Survey. The responses were further analysed based on the type of ticket bought and the geographical distribution. In this case, only passengers travelling on certain ticket types were chosen for further analysis, i.e. seasonal tickets for commuters while passengers travelling using special ticket types and passes were removed from the dataset. This is to ensure better homogeneity of passengers represented in a given demand segmentation category. Having consulted the approach of similar studies and to investigate whether there are any significant differences between passengers, the journey purpose categories were further split by geography (in the case of business and commute) and fare (full and reduced in the case of leisure travellers).

Alongside the proposed demand segmentation, Table 21 provides summary statistics of key variables that are later used in the modelling As can be seen, business and leisure journeys are typically longer and incur lengthier delays, but commuters (who are also more frequent travellers) seem to generally be less satisfied with punctuality levels. This may, to some extent, be explained by commuters being impacted by the largest relative change in journey times due to delays. On average, commuter journey times increase by around 10% due to delays as compared to 5% for other travellers.

Table 21 Demand segmentation summary

	Mean	**SD**	**p25**	**p75**	**p90**	**p95**
BL: Business London (N=18,428)						
Perceived delay (Yes=1)	.21	.41				
Recorded delay (Yes=1)	.52	.50				
Scheduled journey time	80	55	31	117	151	174
Arrival delay	3.59	8.63	0	4	10	16
Departure delay	1.36	4.04	0	1	4	7
Punctuality satisfaction	4.27	.99				
Seated (Yes=1)	.97	.18				
BnL: Business non-London (N=11,251)						
Perceived delay (Yes=1)	.27	.44				
Recorded delay (Yes=1)	.57	.50				
Scheduled journey time	73	57	31	97	158	191
Arrival delay	3.54	7.43	0	4	9	16
Departure delay	2.33	5.52	0	2	6	11
Punctuality satisfaction	4.09	1.12				
Seated (Yes=1)	.93	.25				
CL: Commute London (N=26,020)						
Perceived delay (Yes=1)	.32	.47				
Recorded delay (Yes=1)	.49	.50				
Scheduled journey time	33	19	19	44	58	68
Arrival delay	2.53	5.51	0	3	7	11
Departure delay	1.57	3.94	0	2	4	7
Punctuality satisfaction	3.46	1.30				
Seated (Yes=1)	.78	.41				

CnL: Commute non-London (N=15,216)						
Perceived delay (Yes=1)	.36	.48				
Recorded delay (Yes=1)	.56	.50				
Scheduled journey time	25	16	14	32	46	56
Arrival delay	2.85	5.71	0	3	8	12
Departure delay	2.37	4.67	0	3	7	10
Punctuality satisfaction	3.37	1.33				
Seated (Yes=1)	.80	.40				
LF: Leisure Full (N=24,432)						
Perceived delay (Yes=1)	.18	.38				
Recorded delay (Yes=1)	.45	.50				
Scheduled journey time	50	46	22	60	108	150
Arrival delay	2.36	5.74	0	2	6	11
Departure delay	1.58	4.20	0	1	4	8
Punctuality satisfaction	4.33	.99				
Seated (Yes=1)	.93	.25				
LR: Leisure Reduced (N=51,667)						
Perceived delay (Yes=1)	.21	.41				
Recorded delay (Yes=1)	.51	.50				
Scheduled journey time	85	62	37	119	170	203
Arrival delay	3.45	8.26	0	4	9	16
Departure delay	1.82	5.08	0	2	5	9
Punctuality satisfaction	4.36	.98				
Seated (Yes=1)	.96	.19				

5.5. Summary

This chapter summarised the two sources of data, namely NRPS and HSP and described how the satisfaction and operational data were matched. This will allow studying passengers' perception of delays and the relationship between recorded performance and passenger satisfaction, as introduced later in the thesis. It is worth highlighting the few major limitations related to using NRPS data:

1) While passengers report the origin and destination station they actually travelled between and satisfaction scores are supposed to be based on the specific leg of the

journey, it remains difficult to retrieve the whole journey in the cases where an interchange was needed and establish its potential impact on passenger responses.

2) Linked to the previous point, passengers are asked to score their satisfaction with a specific service they travelled on. However, it remains difficult to know whether the passenger expected to travel on that service or was affected by some cancellations or delays that resulted in the need to travel on a different service, either later or earlier than initially anticipated. Moreover, it is not known if that could ultimately have an impact on the satisfaction score.

3) Similarly, the timing of receiving a questionnaire and completing it may affect the results. Passengers filling in the questionnaire closer to the time of completing the journey may be more likely to have a better recollection of the journey. Similarly, any discontent related to being delayed or poor journey quality may be stronger, leading to more negative survey results.

4) Due to the nature of survey approaches, the NRPS dataset is not necessarily representative of the delay distribution in the network, as the survey design has no control over the distribution of delays experienced by the respondents. However, using data from multiple survey waves makes the NRPS data more resilient to the potential impact of a one-off disruption on the other metrics. Moreover, the NRPS APL levels are in line with network performance estimates provided by ORR.

5) Whilst the number of questions in the survey is relatively large, some of the questions suffer from a relatively low response. This is perhaps a secondary issue for the analysis conducted in this study. Nevertheless, it limits the ability to test the impacts of some of the more specific journey aspects on passenger satisfaction or has an impact on the statistical significance of the estimated results.

6) While some differences between the reported and recorded delays are expected, there is a small proportion of passengers who failed to report very long delays. Whilst it is relatively rare, it may be a result of issues raised under points 1 and 2.

7) The survey consists of a large number of questions, but it might further benefit from adding information about ticket prices and income that would potentially allow calculating metrics such as the value of time from satisfaction data.

8) The nature of delay distribution in the network means that there are relatively few larger delays and it remains more difficult to study the impact of very long delays on passengers. This is further impacted by the larger probability of data errors in

such responses. In the case of this study, however, the main focus is on investigating the impacts of smaller delays.

While the issues highlighted above are worth noting, it is believed that their impact on the modelling undertaken in this study is relatively modest. Nevertheless, some of the highlighted issues may be useful for researchers designing similar surveys in the future.

The choice of variables was carefully described while highlighting the differences between using overall journey satisfaction and satisfaction with punctuality as the key dependent variables. In principle, the main difference between satisfaction with punctuality and overall journey satisfaction is the need to control the latter for satisfaction with different aspects of journey, as delay length is likely to be one of the key, but not the only aspect determining overall journey satisfaction. Whilst there is a large number of variables specifically treating satisfaction with different aspects of the journey, these are often highly correlated and characterised by a low response rate what can become problematic as the modelling strategy effectively discards the responses with missing data.

The final dataset used in the analysis presented in the remaining chapters of the thesis consists of 147,014 responses. The summary of the process of data cleaning is shown in Table 22. While the reduction in the number of responses is sizeable, it still allows for investigating the impacts of delays on a large number of passengers, while reducing the scope for error and improving homogeneity among the studied segments of passengers.

Table 22 Data cleaning and processing steps

Data cleaning step	Number of responses
NRPS full sample	274,862
Responses with a satisfaction score	263,163
Operational data matched	242,311
Delay lengths limit of 30 minutes	240,093
Maximum difference between stated and reported delay within 30 minutes	237,965
Excluded if a delay reported but not matched	224,632
Demand and ticket type segmentation	147,014

The previous chapters provided the motivation behind the thesis, also describing the data used in the remaining chapters that will focus on estimating econometric models of delay perception and satisfaction.

Chapter 6

Rail delays and travellers' perception of being delayed

6.1. Introduction

Transport researchers are often interested in the impact that delays have on passengers and, ultimately, demand and operator revenues. The key aim of a large number of transport infrastructure projects is to reduce travel times with travel time savings being often quantified as the largest benefits of many such investments (for research on the value of time see Mackie et al., 2003; Batley et al., 2019; Schmid et al., 2021). At the same time, a minute of delay has a larger negative impact on travellers and is typically valued at around 3 minutes of scheduled journey time (see Preston et al., 2009; Wardman and Batley, 2014; Wardman and Batley, 2022 for reviews on lateness valuation).

When choosing a service to travel on, passengers consider the scheduled arrival time of the services with respect to their preferred arrival time. As noted by Preston et al. (2009), explained in more detail in Batley (2007) and discussed in Chapter 2, passengers typically also include a safety buffer to their travel schedules that works as a time insurance against any possible disruptions and aims to increase the probability of arriving to the destination within passenger's preferred arrival time. The amount of buffer time depends on expectations, sensitivity to potential lateness, and personal characteristics of a passenger.

While intuition suggests that performance affects levels of demand, passengers are not always able to change their travel behaviour following late running, at least in the short term (Batley et al., 2011). This means that while delays may have a negative impact on passengers, performance may not always be immediately linked to demand and an alternative way to study the delay impacts on passengers is to look at the relationship between delays and journey satisfaction. Several studies have looked at the impact of different aspects of the journey on travel satisfaction (for reviews see De Vos et al., 2013; De Oña and De Oña, 2015; Gao et al., 2018; Rong et al., 2022). Travel time, prices, journey comfort and provision of information are of paramount importance for travellers (Dziekan and Kottenhoff, 2007; Brons and Rietveld, 2009; St-Louis et al., 2014; Susilo and Cats, 2014; Mouwen, 2015; De Oña et al., 2016; Abenoza et al., 2019; Monsuur et al., 2021).

Previous research suggests that passengers delayed by over 30 minutes are very unlikely to be satisfied with their journeys (Wittmer and Laesser, 2010; Monsuur et al., 2021). Moreover, according to Monsuur et al. (2021), in the case of standing passengers, this threshold reduces to 10-20 minutes. Nevertheless, as noted by Transport Focus (2015), satisfaction levels tend to start dropping from the very first minute of late running. For business and leisure travellers they decrease somewhat less rapidly until a threshold of respectively 5 and 8 minutes of lateness is reached.

A large body of literature is devoted to the valuation of time changes, both in terms of reducing scheduled journey times and increasing punctuality and reliability (e.g. Mackie, et al., 2003; Preston et al., 2009; Batley and Ibáñez, 2012; Batley et al., 2019; Schmid et al., 2021). Mackie et al. (2003) and Daly et al. (2014) discussed the impacts of small time savings whereas Daly et al. (2014) noted that there is no consensus on the treatment of benefits arising from these small changes with differences in standard practices around the globe. As noted by Welch and Williams (1997), minor savings often account for a large proportion of the time benefits of transport projects. However, as suggested by Daly et al. (2014), it is likely that assuming a constant marginal value of time is wrong as passengers might not be able to notice the smaller journey time reductions. Though, Mackie et al. (2001) argued that even if passengers are not able to perceive these changes, they can still have economic benefits, similarly to the impacts from investment in road safety where changes in accident probabilities are not directly observed or noticed by the road users. Analogously, smaller delays that do not necessarily affect passenger satisfaction, or even remain unperceived, are not necessarily unimportant. Increased understanding of the impacts of smaller delays can contribute to exploring any potential inherent non-linearities in the valuation of time.

While previous research provides us with some understanding of how passenger satisfaction changes with delays, little is known about how passengers perceive delays and the effects that delay perception may have on satisfaction. Nielsen (2000) and Rezapour and Ferraro (2021) indicated that passenger perception of late running has an impact on travel behaviour and public transport suppliers can learn how to improve their services by investigating these impacts too.

In principle, work conducted as part of this chapter draws on a number of previous studies examining the impacts of lateness on passengers (i.e. Batley, 2007; Preston et al., 2009; Monsuur et al., 2021) whilst responding to the suggestion in Wardman and Batley (2022) and Rong et al. (2022) that looking at the differences between perceptions of late time and recorded delay lengths can improve understanding of the impact that delays have on passengers. In this context, the perception of delay is thought of as an intermediate step linking the existence of delay (supply-side disruption) with the impacts on utility and ultimately demand (demand-side impact), since for delays to have an impact on travellers, they clearly need to be perceived.

The work conducted as part of this chapter uses data on rail passenger delay perception from 10 waves of a travel satisfaction survey (NRPS) conducted by Transport Focus in the United Kingdom. Passenger responses were matched to the operational data using the Historic Service Performance (HSP) database to allow comparison between passengers' reports of late arrival and recorded delay lengths as described in Chapter 5. The investigation of delay perception is based on the analysis of:

1) travellers' ability to perceive delays (i.e. Q1 from Table 1)
2) passengers' misperception or rounding of reported delay lengths (i.e. Q2 from Table 1) and
3) the impact of journey type, length, comfort as well as arrival versus departure delay on how delays are perceived (i.e. Q3 from Table 1)

First, binary response models are estimated where passengers' ability to perceive a delay is explained by delay length (both at the departure and arrival) while also controlling for journey length, quality, and type. The additional controls act as hypothesis tests to investigate whether they have a significant impact on how delays are perceived. Subsequently, additional analysis is conducted to better understand the possible misperceptions of delay lengths.

The remainder of this chapter is structured as follows:

- Section 6.2 provides a review of literature covering human perception in a variety of fields to describe the main differences in the ways people perceive the same stimulus and methods used to validate perception.
- Section 6.3 discusses the possible determinants of delay perception.
- Section 6.4 summarises how the NRPS dataset can be applied to analysing delay perception and reports the results of the analysis, first looking at delay perception as a binary outcome (i.e. delay is perceived or not).
- Section 6.5 investigates the possible reasons as to why delays were reported despite on-time performance.
- Section 6.6 looks at the relationship between perceived and recorded delay lengths.
- Section 6.7 provides a summary of findings and conclusions.

6.2. Review of studies on human perception

Human perception plays an important role in behaviour and decision-making. As noted by Manski (2004), Shepperd et al. (2013) and Lupyan (2017), the knowledge of survey respondents may be partial, leading to discrepancies between self-reports or expectations and actual values. To better understand how the perception of delays differs from actual performance, a review of studies on human perception was conducted across different disciplines. There is an abundance of literature looking at human perception in different contexts, which differ with respect to how easy it is to observe and measure a stimulus and subsequently compare the perception to measured/observed values. Table 23 below provides a multidisciplinary perspective on human perception.

Having reviewed research on human perception in various contexts, it is now of interest to discuss the likely differences and similarities to delay perception. Crime perception research offers some insights into the divergence between perception and reality, but unlike transport delays, the crime statistics are not directly experienced by those reporting safety concerns. On the other hand, pain perception is conceptually closer to transport delays as pain is, indeed, directly experienced. However, pain levels are subjective as significant heterogeneities in their reporting have been found across patients. This means, that unlike delays, which have an objective measure (time), providing an objective measure of pain is not always easy. Similarly, while the probability of catching a virus can be calculated, the risk itself is not directly experienced as the outcome here is binary – either becoming ill or not. Walkability and accessibility, despite being transport-related concepts, are also difficult to be objectively quantified, i.e. there is no single measure that can capture them, since using distance as a proxy is argued to be an oversimplification. In terms of measurability, a comparison of perceived and actual test performance is probably most similar to train delays. However, significant conceptual differences remain - in the case of delays, the outcome is binary as is in the case of catching a virus. Though, a delay may need to cross some threshold for passengers to be able to notice it (this might depend on lateness sensitivities or tolerance) and is conceptually more similar to pain perception, in the sense that the longer the delay, the larger the negative impact and, hence, the probability of perceiving it.

Table 23 Literature review of studies on human perception

Measure	Observability	Perception	References
Crime rates	Measurable, not directly observable	Tendency to overestimate. Influenced by personal views, news, and word of mouth. Perception has a larger impact on behaviour than the crime rates. Subject to biases such as feeling that familiar areas are safer.	Spicer et al., 2014; Lora, 2016; Vallejo Velazquez et al., 2020; Manning et al., 2022
Probability of being caught committing a crime	Measurable, not directly observable	Perception of the probability of being captured after committing a crime affects crime rates. In an educational setting, the perceived risk of punishment has an impact on self-reported delinquency in students.	Bailey et al., 1974; Jensen et al., 1978; Lochner, 2007
Pain perception	Pain is a stimulus, but not directly measurable or comparable	Sense of touch is not the only one playing a role in experiencing pain. Pain reporting methods used in medical research and clinical applications include descriptive, verbal, numerical scales, and pain drawings. Personal characteristics have an impact on perception and reporting.	Freyd, 1923; Budzynski et al., 1973; Margolis et al., 1986; Crombez et al., 2005; Haefeli and Elfering, 2006; Mancini et al., 2011; Verma et al., 2015; Cimpean and David, 2019; González-Roldán et al., 2020; McIntyre et al., 2020
Test performance	Results are measurable, but observable after the performance is perceived	Significant differences between perception and actual test performance. Overconfidence of those who performed poorly.	Ehrlinger et al., 2008; Papamitsiou and Economides, 2014
Risk of catching viral diseases	Measurable, but difficult to calculate for all the activities	Reported infection rates and crowding levels determine perceived risk levels and influence behaviour.	Lau et al., 2003; Boes and Winkelmann, 2006; Schneider et al., 2021; Lewis and Duch, 2021; Shelat et al., 2022; Cipolletta et al., 2022
Accessibility and walkability	Measurable, but no single metric able to describe it	Calculated measures can only serve as a proxy for perceived accessibility.	Saelens et al., 2003; Frank et al., 2010; Lättman et al., 2016; Tiznado-Aitken et al., 2020; Pot et al., 2021; De Vos et al., 2023

In transport, there has been very limited research concerning travellers' perceptions, especially in the context of delays. A large number of studies looked at the impact of various aspects of the journey on traveller satisfaction (e.g. Brons and Rietveld, 2009; De Vos et al., 2013; Susilo and Cats, 2014; Cats et al., 2015; Mouwen, 2015; De Oña and De Oña, 2015; De Oña et al., 2016; Machado-León et al., 2017; Gao et al., 2018; Obsie et al., 2020; Monsuur et al., 2021; Rong et al., 2022). However, in most cases, it is difficult to validate perception with a quantifiable service quality measure (Eboli and Mazzulla, 2021). For example, satisfaction with seat comfort may depend on the type of seat, but the type of seat is a categorical variable and it cannot be immediately expressed by a numeric value, as is the case with time or money. On the other hand, delays have an objective quantifiable measure - time. In this sense, validating the perception of delays can help understand the relationship between the objective delay (actual performance), subjective delay (perception) as shown in Figure 30, and subsequently travel satisfaction (utility) or demand.

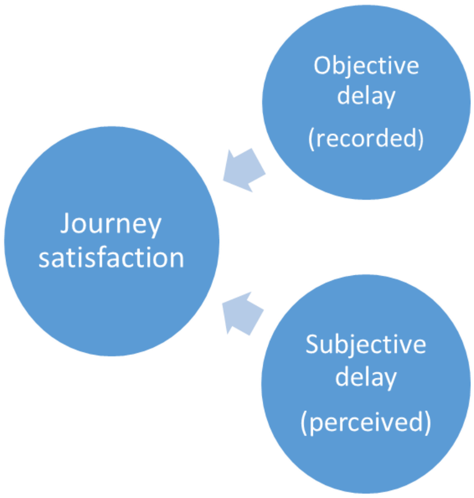


Figure 30 Objective and subjective delays

Whilst investigating the impacts of delays on passenger satisfaction, most studies typically focus on the relationship between perceived lateness and journey satisfaction. As suggested by Friman and Fellesson (2009), objective (actual) performance does not always perfectly explain passenger satisfaction. However, as noted by De Oña and De Oña (2015) and Eboli and Mazzulla (2021), objective performance indicators are more useful than the subjective ones as they are unbiased. As journey satisfaction data typically come from passenger surveys, the journey lengths and delays are rarely validated with operational data. As argued by Nathanail (2008) and Rong et al., (2022), combining perceived delays with recorded delays can improve understanding of the impacts of lateness on passenger satisfaction. A few studies where this is explored include:

- Higgins et al. (2018) looking at the relationship between satisfaction and duration of car commute combined with respondents' perceptions of congestion levels in Canada. As suggested by this study, drivers who perceive congestion as a serious and frequent issue tend to be less satisfied with their travel times.
- Carrel et al. (2016) matched transit passenger satisfaction data with smartphone location data from the San Francisco Travel Quality Study to relate satisfaction to unreliability levels and estimated a 69-percentage-point decrease in the proportion of satisfied bus passengers following a 10-minute delay.
- Gao et al. (2018) estimated the impacts of differences between expected and experienced access, egress, and in-vehicle time on the satisfaction of public transport users in the Chinese city of Xi'an, suggesting that the gap between the two can better explain satisfaction levels than the absolute values.
- Using data from the Greek rail national survey in combination with operational data, Nathanail (2008) compared performance and satisfaction levels and suggested that improvements in station and train facilities as well as developments of passenger information systems were of paramount importance for Greek railways at that time.
- Rong et al. (2022) studied the relationship between actual and perceived performance and passenger satisfaction of bus users in the Chinese city of Shijiazhuang, concluding that actual travel time is not the most important factor influencing passengers' perception of travel time. Instead, they suggested that the number of stops and stopping times are more important, arguing that these counter-intuitive results highlight passengers' inability to perceive time accurately. The perception of travel time was suggested to be largely influenced by the negative emotions related to judgment impacted by stopping and speed. While there are inherent operational differences between buses and trains, this highlights the importance of factors other than arrival delay in determining perception of delays or their consequential satisfaction impacts.

The review of the literature on the perception in very different contexts leads to formulation of the following hypotheses related to how delays may be perceived, given that delay has an objective and quantifiable measure, time:

1) It may be expected that the length of delay has to reach a certain threshold to be perceived.

2) The outcome is binary – a delay is either perceived or not. However, the outcome will be continuous in the case of looking at the perception of delay lengths.

3) The abundance of different biases affecting perception may also affect delay perception, e.g. travelling in comfortable versus crowded conditions may have an impact on the ability to perceive delays due to the negative emotions related to low journey comfort.

The following section discusses the likely determinants of delay perception as it is noted that the length of delay is likely not the only factor affecting delay perception.

6.3. Determinants of delay perception

Before discussing the determinants of delay perception and the methodological approach, it is worth defining some key terms as shown in Figure 31 below:

- Scheduled departure and arrival times are the scheduled times where the train was due to depart from the origin station and arrive at the destination with the difference between scheduled arrival time and scheduled departure time equalling to scheduled journey time.

- Departure delay and arrival delay are the differences between actual and scheduled departure and arrival times with actual journey time being calculated as the difference between actual arrival at the destination and scheduled departure time.

As generally suggested by the perception research, perception of a given stimulus or phenomenon is typically not only dependent on its intensity but also on personal characteristics or experiences. Conceptually, delay perception may be similar to pain perception where the research interest lies in understanding the pain levels (or thresholds) that start being noticeable by patients. In the case of delay perception where the focus is on understanding how the ability to perceive delays changes with increasing recorded delay lengths, this can be understood as determining the delay length where passengers start noticing that they arrived late. While the intuition suggests that the length of delay at arrival is the key motivator of delay perception, it is of interest to investigate if there are any other factors affecting how delays are perceived.

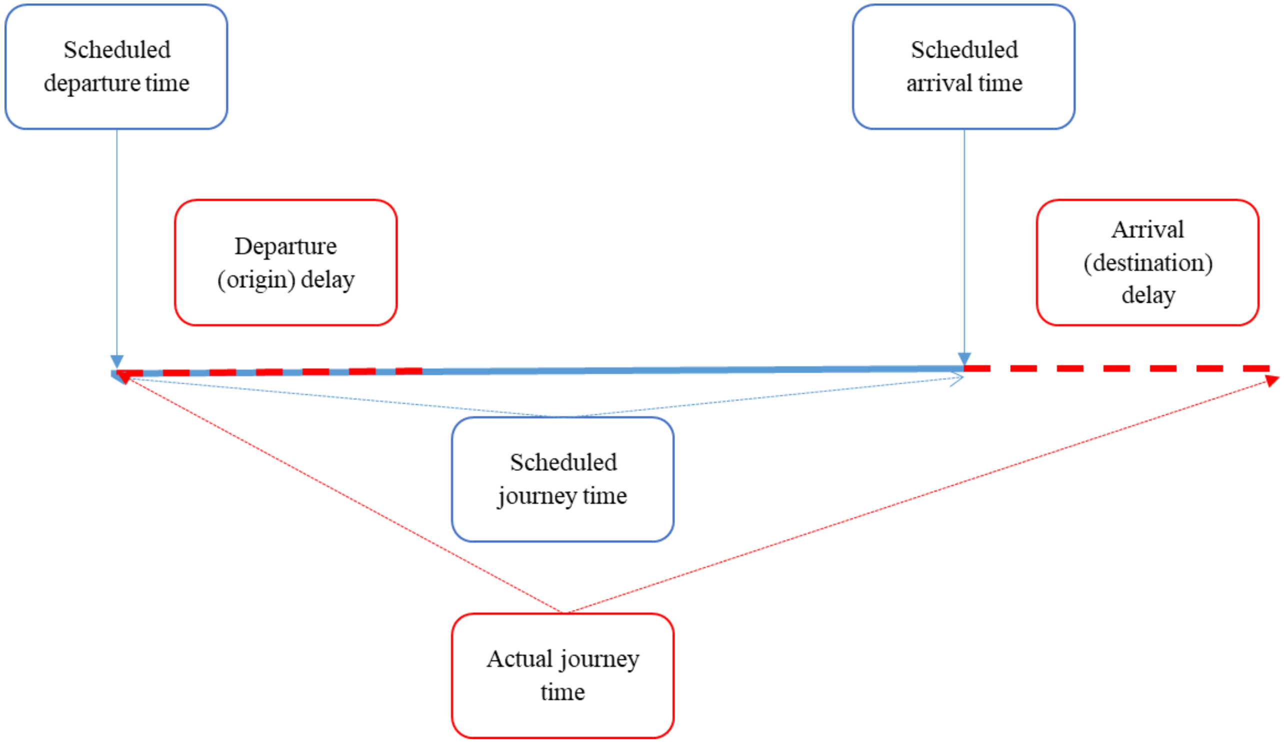


Figure 31 Definitions of delay and journey time

Whilst satisfaction effects are described in more detail in Chapter 7, it is worth highlighting the key findings from the literature on passenger satisfaction as journey aspects affecting satisfaction may, indeed, also affect delay perception. Nonetheless, the length of the experienced delay is expected to be the key determinant of delay perception with other factors expected to have an indirect (complementary) effect on how the experienced delays are perceived. These characteristics are assumed to only affect delay perception if the delay occurred, either increasing or decreasing the ability to perceive a given delay length. Figure 32 and the paragraphs below describe these expected complementary effects.

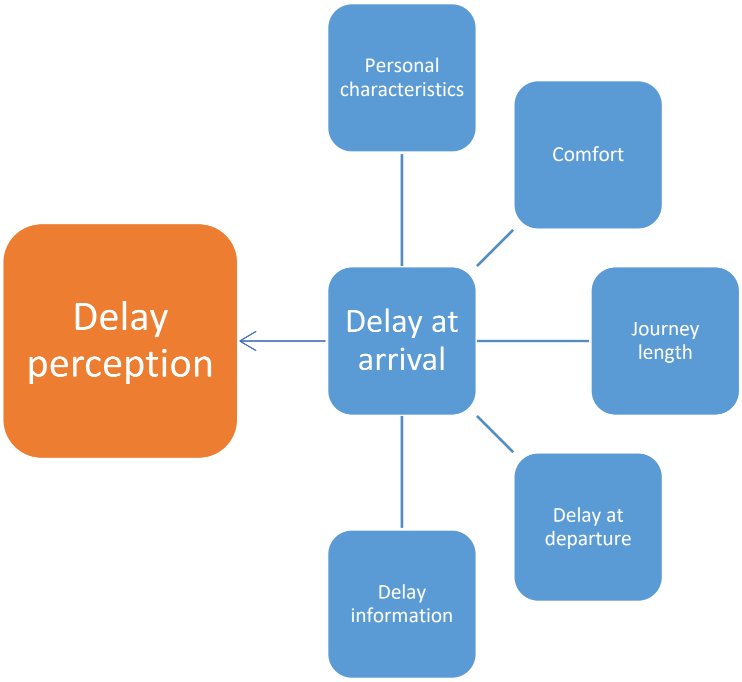


Figure 32 Determinants of delay perception

Personal characteristics

In the case of pain, personal characteristics were suggested to play a very important role in how it is perceived. Passengers who are more sensitive to delays (i.e. commuters) may be more likely to perceive delays due to a larger focus on on-time arrival resulting from the inclusion of smaller buffer times and/or larger negative consequences of late arrival (i.e. being late for work) or better knowledge of timetables.

Comfort

Previous research also indicated the important role of journey comfort as increasingly, passengers try to find ways to use their travel time productively (either working or focusing on leisure activities) (Lyons and Urry, 2005; Lyons et al., 2007). Similarly, the valuation of journey time is typically higher for passengers who travel in crowded conditions as compared to more comfortable settings. Seated passengers may use their travel time more productively, in turn, being less concerned with performance.

Journey length

It can be expected that passengers' ability to perceive a delay may differ between shorter and longer journeys and is likely to decrease with journey length. Therefore, passengers who travel on shorter journeys may be more likely to notice a smaller difference in travel time as compared to passengers travelling on longer journeys. This may be related to the inclusion of different safety margins and a smaller delay already representing a large increase in journey time (as compared to longer journeys).

Delay at departure

While from the operational point of view, the main focus is on the length of delay at arrival (destination), it may be expected that from the passenger's point of view, the perception of performance is formed by personal beliefs, judgments and/or experiences. Previous research, particularly the work by Batley and Ibáñez (2012), found significant differences in the impact of departure versus arrival delays on passengers. It may be expected that a delay at boarding may distort delay perception due to the additional stress and discomfort related to uncertainty and waiting for a delayed train on platform. As passengers do not necessarily know the exact scheduled departure and arrival times (Rietveld, 2002), a large delay at departure may affect the perception of final performance.

Delay information

Similarly, related to personal beliefs or experiences, passengers' ability to perceive a delay may also increase if real-time information is provided on-board or announcements about the delays are made to passengers at stations.

The following sections describe the methodological approaches and report the results of the analysis.

6.4. Binary outcome models of delay perception

6.4.1. Data

In investigating delay perception, the analysis introduced in this chapter employs the data from NRPS. Passenger responses recorded as part of ten NRPS survey waves between 2015 and 2020 were matched to operational data using the HSP database as described in section 5.2. This allowed computing scheduled journey length, actual journey length, and recorded delays (at departure and arrival) for each of the passengers taking part in the survey. The recorded delays were compared with passengers' reports about their experience of delay (Figure 26).

The introductory analysis of delay perception was presented in section 5.2.2 and highlighted that the proportion of travellers reporting being delayed in the sample increases with the length of recorded delay. The subset of the data used in the main body of analysis corresponds to the dataset summarised in Table 21, but the analysis of delay perception generally focuses on the responses with a matched delay. However, smaller subsets of the dataset are used in the cases where some variables of interests were not included in all the analysed survey waves.

The analysis conducted as part of this chapter looks at:

- passengers' general ability to perceive a delay (section 6.4),
- reported delays when on-time performance was recorded (section 6.5) and
- the perception of delay lengths (section 6.6).

For the first two streams of analysis, binary outcome models of delay perception are estimated whilst the latter is explored through analysing correlations between reported and recorded delay lengths.

The following section introduces the modelling approach undertaken to study passengers' general ability to perceive a delay that represents the main stream of the analysis presented in this chapter. The alternative lines of investigation are presented in the relevant sections (6.5 and 6.6).

6.4.2. Methodology

With the aim of analysing the relationship between the recorded length of delay and passengers' perception of being late, the aforementioned delay reports were converted to a binary response variable with a passenger either reporting late arrival or arriving on time:

$$Y = \begin{cases} 1\ if\ passenger\ reported\ late\ arrival \\ 0\ if\ passenger\ reported\ arriving\ on\ time \end{cases}$$

(8)

To increase the understanding of the impact of delays on passengers, this work aims to investigate how the probability of reporting a delay changes with increasing delay lengths. This type of modelling requires the usage of binary outcome modelling methods. In this case, the modelling approach allows estimating the probability of a passenger reporting late arrival (i.e. $Y = 1$) versus arriving on time (i.e. $Y = 0$) with the delay perception defined as:

$$P(Y = 1) = F(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)$$

(9)

where

$$F = \frac{1}{(1 + e^{-(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)})}$$

(10)

and X_i is an explanatory variable i, and β_i is the corresponding parameter.

If F is the cumulative normal distribution, then the binary response model is referred to as probit. If instead, it follows the logistic distribution, then the model is referred to as logit. Both distributions are symmetrical and the main difference is that the logistic distribution has longer tails (Horowitz and Savin, 2001).

The binary outcome models were estimated using the logit model function in Stata 17 (StataCorp, 2021). The estimated model coefficients represent the rate of change in the log-odds as the estimated model has the following form:

$$\log(\frac{p}{1-p}) = \beta_o + \beta_1 X_1 + \cdots + \beta_i X_i$$

(11)

and

$$p = \frac{1}{(1 + e^{-(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)})}$$

(12)

The estimates can be indicative of the direction of the relationship and size but are relatively difficult to interpret directly. In addition to the estimated coefficients, estimated probabilities are shown graphically. Moreover, the delay length thresholds for the probability of perceiving a delay of 0.5 (binary response cut-off where probability smaller than 0.5 is counted as 0 and probability of 0.5 or more is counted as 1) are compared for the different types of models estimated for comparison purposes, i.e.:

$$0.5 = \frac{1}{(1 + e^{-(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)})}$$

(13)

The so-called ‘delay thresholds’ are estimated using the model results as this presents a way to indicate a critical level needed to change the outcome, what is easily translated for policymaking. Moreover, this demonstrates a useful way of comparison between the different types of travellers as well as when contrasting the lengths of delays that are perceivable versus those that have a detrimental impact on delay satisfaction (as introduced in Chapter 7). This approach is used throughout the thesis and the forecasted thresholds are generally presented alongside the forecasted outcomes shown in plots that depict the shape of the relationship in more detail. Average marginal effects are not generally reported due to the limited merit when comparing between the impacts of delays on delay perception and satisfaction. This is primarily driven by the fact that under the ‘no delay’ scenario, the probabilities of being satisfied are not necessarily equal to 0 what will be further discussed in Chapter 7.

Four models were considered with delay perception explained by delay length at arrival in the initial model and extended models with inclusion of the additional variables as shown in Table 24.

Table 24 Description of the estimated models

Measure	Variable	Expected effect on delay perception (hypothesis tested)	Model 0	Model 1	Model 2	Model 3
Delay at arrival	Recorded length of arrival delay (L_A)	Positive	X	X	X	X
Personal characteristics	Journey purpose categorisation (*JP*)	Commuters more likely to perceive delays		X	X	X
Comfort	Passengers’ reports about seat availability (*Seat*)	Negative, seated passengers less likely to notice delays				X
Journey length	Scheduled journey time (*SJT*)	Negative, long-distance travellers less likely to notice delays			X	X
Delay at departure	Recorded length of delay at departure (L_D)	Positive, departure delay affecting perception of final performance				X
Delay information	Not represented	Positive, information increasing passengers’ focus on delays				

The initial model (model 0) has the following form with delay perception modelled as a function of arrival delay for responses where a delay at arrival was matched:

$$\log(\frac{p}{1-p}) = \sum_{i=1}^{i=6}(\beta_1 + \beta_2 \times L_A^+)$$

(14)

In model 1, model 0 is extended to allow for heterogeneity due to journey purpose, i.e.:

$$\log(\frac{p}{1-p}) = \sum_{i=1}^{i=6} (JP_i \times \beta_{1,i} + JP_i \times \beta_{2,i} \times L_{A,i}^{+})$$

(15)

where:

JP_i is a journey purpose dummy variable for each of the 6 journey purposes that takes the value of 1 when it matches the respondents' journey purpose or 0 otherwise

$L_{A,i}^{+}$ is the delay length at arrival which is defined as the difference between the actual and scheduled arrival for all cases where the difference is positive; when the difference is negative, such responses are treated as on-time arrival

Model 2 aims to test the hypothesis that the marginal probability of perceiving a delay differs with journey lengths. An interaction term between scheduled journey length, arrival delay and journey purpose was introduced to the model. Testing how the impacts vary with journey purpose and lengths is in line with typical rail economics research (i.e. Batley and Ibáñez, 2012). However, journey length enters the model as a continuous, rather than categorical variable, which is often the case when flows are segmented into journey length categories in SP studies (i.e. Batley et al., 2019). In SP studies with just a few different flow types, it may not always be practical to use continuous variables due to the relatively few distinct values. Hence, the model takes the following form:

$$\log\left(\frac{p}{1-p}\right) = \sum_{i=1}^{i=6} (JP_i \times \beta_{1,i} + JP_i \times \beta_{2,i} \times L_{A,i}^{+} + JP_i \times \beta_{3,i} \times L_{A,i}^{+} \times SJT_i)$$

(16)

where:

SJT_i is the scheduled journey time

Model 3 extends the previous model by investigating the impacts of arrival versus departure delay as well as journey quality on delay perception. This is to represent the possible complementary nature of departure delay's impact on the perception of arrival delay. This is done by introducing an interaction term between arrival and departure delay to the model. The arrival delay coefficients are also estimated separately for standing and seated passengers, with the seat dummy variable being a proxy for journey quality and taking the value of 1 if passenger was able to find a seat or 0 otherwise. As passengers' reports of seat availability were only available for 7 out of the 10 survey waves, the sample size used for model 3 is smaller than for the simpler models. This model takes the following form:

$$\log\left(\frac{p}{1-p}\right) = \sum_{i=1}^{i=6}(JP_i \times \beta_{1,i} + Seat_i \times JP_i \times \beta_{2,i} \times L_{A,i}^{+} + Seat_i \times JP_i \times \beta_{3,i} \times L_{A,i}^{+} \times SJT_i + JP_i \times \beta_{4,i} \times L_{A,i}^{+} \times L_{D,i}^{+}) \quad (17)$$

where:

$Seat_i$ is a dummy variable that takes the value of 1 if passenger reported having a seat or 0 otherwise

$L_{D,i}^{+}$ is the delay length at departure which is defined as the difference between the actual and scheduled departure. If a train departed before its scheduled departure time, this is counted as on time departure.

Additional sensitivity analysis is introduced in Annex I.

It is noted that the proposed models do not include all levels of interacted variables. This was done on purpose with the aim of only capturing the complementary nature of some of the explanatory variables. The main effect is then captured by the length of delay at arrival and if any other explanatory variables are introduced, they are introduced as an interaction between the length of delay at arrival and that variable. The explanatory variables included scheduled journey length, length of delay at departure and a dummy variable representing whether a passenger was seated or standing. For completeness, the fully specified models are also estimated (this applies to models 2 and 3) and compared to the proposed models.

6.4.3. Results

As noted in the previous sections, the ability to perceive a delay is modelled as a binary outcome based on respondents' reports of late running. The probabilities of reporting a delay are estimated and the delay perception thresholds are predicted for increasing recorded lengths of delays for different types of journeys, also allowing investigation of the impact of journey quality, length and arrival versus departure delay on delay perception. The results of the simpler versions of the model are introduced first, followed by the extended version of the model (i.e. model 3) with more explanatory variables, with additional sensitivity analysis presented in Annex I.

Initial models of delay perception

This section reports the results of the binary outcome models of delay perception introduced earlier in section 6.4. The following models are introduced first with the results presented in Table 25:

- Model 0 being the starting point of the analysis where delay perception is explained by the length of delay at arrival.
- Model 1 that allows for heterogeneity due to different journey purposes.

- Model 2 investigating the complementary impact of scheduled journey length.
- Model 2a that includes all levels of the interacted variables from model 2.

In line with expectations, the arrival delay coefficients are positive and statistically significant in models 0 and 1 as the probability of perceiving a delay increases with delay length. Introducing journey purpose segmentation (in model 1) improves R^2 and a likelihood ratio test (Wilks, 1938) was run to determine the significance of the increase in the explanatory power, generally justifying the added complexity (LR test statistic of 2584.59, p-value of less than 0.0001). Figure 33 shows the predicted probabilities of reporting late running for each of the journey purposes and increasing delay lengths. The predicted delay length thresholds at 0.5 probability cut-off (p=0.5) suggest that commuters become more likely to perceive a delay after arriving around 4-5 minutes late. For other types of passengers, this delay threshold is between 8-11 minutes. On average, a minute of delay increases the probability of noticing it by around 0.025-0.03 for business travellers to London, 0.03-0.035 for leisure travellers, 0.035 for non-London business travellers, 0.04 for commuters to London and 0.045-0.055 for non-London commuters. It is worth highlighting, that for commuters, the probability of delay perception is suggested to reach almost 1 at delay lengths of 15 minutes. For the other travellers, the corresponding probability is around 0.75. This would suggest that commuters are more sensitive to delays, resulting from inclusion of shorter buffers around their PAT or the fact that commuter journeys are typically shorter (see Table 21) and a smaller delay already represents a significant proportional change in journey time.

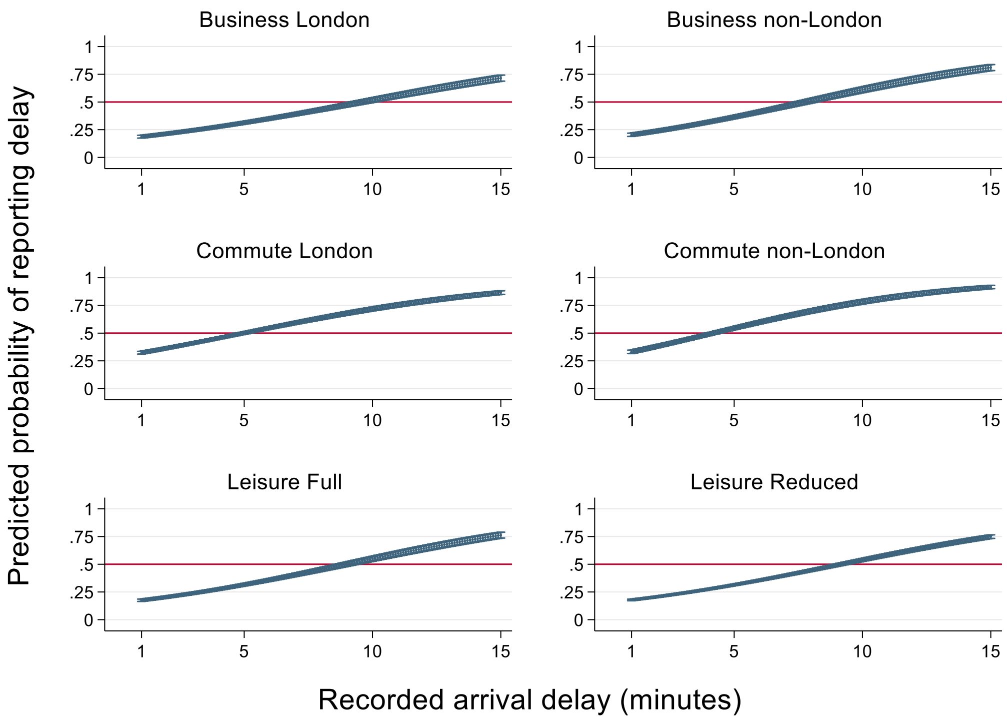


Figure 33 Estimated probabilities of delay perception under model 1 from Table 25

Table 25 Estimated coefficients from the binary logit models of delay perception (models 0-2a)

	(0)	(1)	(2)	(2a)
Constant	-1.437***	-1.741***	-1.784***	-1.744***
	(-114.44)	(-46.65)	(-46.70)	(-25.64)
BnL		0.215***	0.187**	0.357***
		(3.72)	(3.15)	(3.55)
CL		0.714***	0.730***	0.695***
		(15.06)	(15.07)	(7.75)
CnL		0.822***	0.823***	0.893***
		(15.62)	(15.32)	(9.29)
LF		-0.0637	-0.0854	0.0355
		(-1.25)	(-1.63)	(0.41)
LR		-0.0148	-0.0463	0.252**
		(-0.34)	(-1.03)	(3.22)
Arrival delay				
BL	0.166***	0.153***	0.195***	0.191***
	(86.22)	(32.11)	(23.63)	(19.52)
BnL		0.185***	0.259***	0.237***
		(26.92)	(24.51)	(19.43)
CL		0.187***	0.238***	0.237***
		(33.28)	(25.84)	(20.95)
CnL		0.225***	0.305***	0.290***
		(28.11)	(24.90)	(20.01)
LF		0.185***	0.241***	0.225***
		(33.69)	(31.07)	(26.09)
LR		0.171***	0.239***	0.208***
		(55.70)	(47.80)	(36.44)
Arrival delay x SJT				
BL			-0.000358***	-0.000316***
			(-6.50)	(-3.93)
BnL			-0.000726***	-0.000458***
			(-10.47)	(-4.42)
CL			-0.00122***	-0.00120***
			(-7.57)	(-5.00)
CnL			-0.00249***	-0.00194***
			(-9.59)	(-5.08)
LF			-0.000697***	-0.000419***
			(-11.24)	(-4.47)
LR			-0.000541***	-0.000213***
			(-18.74)	(-5.00)
SJT				
BL				-0.000492
				(-0.72)
BnL				-0.00285***
				(-3.50)
CL				-0.000147
				(-0.10)
CnL				-0.00427
				(-1.93)
LF				-0.00326***
				(-4.00)
LR				-0.00398***
				(-10.42)
N	73050	73050	72884	72884
LL	-42557.5	-41265.2	-40772.8	-40699.0
Pseudo R^2	0.107	0.135	0.143	0.144

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; ^ difference between seated and non-seated coefficients being statistically significant; BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced.

Model 2 extends model 1 by the addition of an interaction term between scheduled journey time and arrival delay to test the hypothesis that passengers travelling longer would need to be subjected to a lengthier delay to perceive it. The estimated coefficients are presented in Table 25 above (under model 2). As expected, the probability of perceiving a delay is suggested to increase with increasing delay length. However, given the negative coefficients on the interaction between delay length and scheduled journey time, the probability of perceiving a delay is suggested to decrease with journey length. Figure 34 below shows the predicted probabilities of perceiving a delay for all the journey purposes for increasing delays at the 10^{th} and 90^{th} percentile of scheduled journey time. This is to take into account the differing distribution of journey lengths. As the predicted probabilities show, typically for a longer journey, a longer delay is needed to achieve the same probability of delay perception. The delay perception thresholds where the predicted probability is 0.5 are shown in Figure 35 below for the 10^{th} and 90^{th} percentiles of scheduled journey lengths. Therefore, the change in the journey length from the 10^{th} to 90^{th} percentile increases the delay length threshold required for a passenger to be more likely to perceive a delay. The difference between the delay length thresholds for journeys at the 10^{th} and 90^{th} percentile of journey length is smallest for commuters (1.7 minutes) followed by business travellers to London (3.2 minutes), leisure travellers on the full fare (3.5 minutes), non-London business travellers (4.6 minutes) and leisure travellers on reduced fares (4.7 minutes). However, both the estimated thresholds and scheduled journey lengths differ between the demand segments. It can be seen that at the 10^{th} percentile of scheduled journey length, the delay length thresholds show considerable variation while scheduled journey lengths are similar, as the delay length threshold is the lowest for non-London commuters at 3.4 minutes for a journey time of 22.5 minutes compared to 9.6 minutes for a journey time of 22 minutes for business travellers to London. For the longer journeys, the change in the delay thresholds seems to be driven by scheduled journey lengths rather than journey purpose. Overall, model 2 would suggest that some of the differences estimated in model 1 are, in fact, partly explained by the inherent differences in journey lengths between the different passenger categories.

Model 2a expands the previous specification (of model 2) by incorporating scheduled journey length on its own (in addition to its interaction with delay length). However, as shown in more detail in Annex II (section A), there are no significant differences between the results of the two estimated models. While it is generally standard practice to include all levels of the interacted variables, the main interest here is on the impact that arrival delay has on the perception delay and the complementary nature of scheduled journey time. Conceptually, scheduled journey time itself should not impact the probability of perceiving a delay independently of delay. If it does affect it directly (i.e. as allowed in model 2a), then this could be considered to be an irrational response, unless this demonstrates the impacts of the expectations about delays on different journey types. Hence, the preferred model in this case remains model 2.

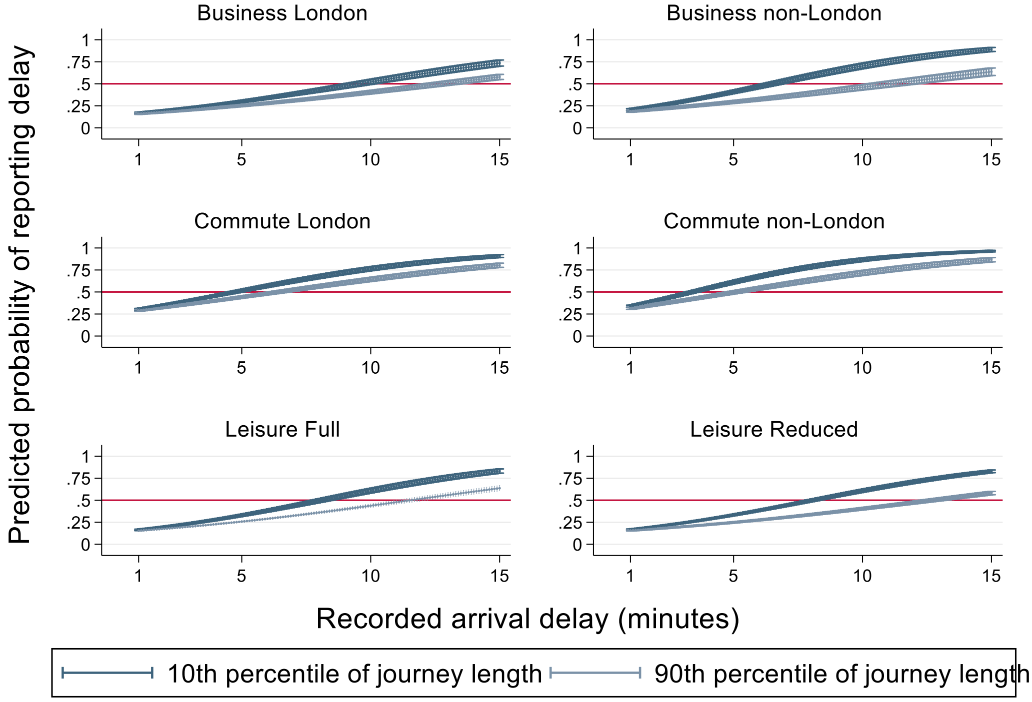


Figure 34 Estimated probabilities of delay perception under Model 2 from Table 25

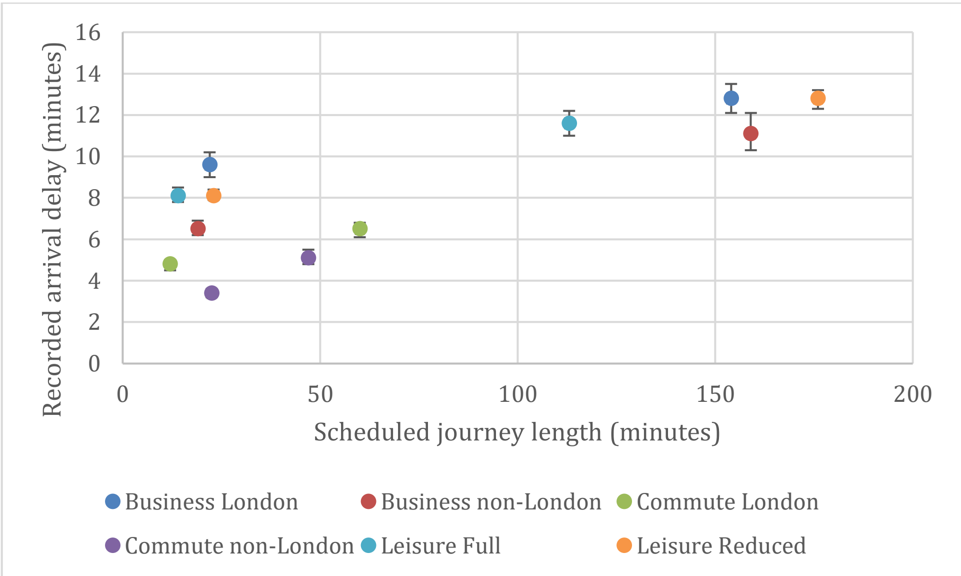


Figure 35 Estimated delay thresholds for p=0.5 and 10th and 90th percentile of SJT

Extended models of delay perception

Subsequently, the delay perception is expected to be additionally impacted by journey comfort, onboard announcements or length of delay at departure as:

1) passengers may be more likely to notice a delay when journey comfort is worse,

2) passengers waiting for a delayed train at departure may be more likely to perceive a delay than passengers who were delayed while being on-board or

3) on-board and platform announcements may influence passengers' perception (though this is not explored as part of this analysis due to lack of data).

Against this background, model 3 extends model 2 by estimating the coefficients separately for passengers who reported being seated and standing with results reported in Table 26. Additionally, an interaction between arrival and departure delay is added to represent the impact of the length of delay at the departure on delay perception for a given length of arrival (destination) delay. Not including all levels of interacted variables in model 3 is deliberate, aiming to focus on testing the complementary effects of the additional explanatory variables to the main effect of the length of arrival delay on delay perception. However, for completeness and to allow comparison, model 3 is also re-estimated with inclusion of all levels of interacted variables (model 3a in Table 26).

Model 3 includes a larger number of explanatory variables, but due to data limitations, the sample size is reduced when compared to the previous models. Hence, assessing whether the more complex model significantly improves the fit using a likelihood ratio test (Wilks, 1938) is complicated. However, such a test was run following re-estimation of model 1 with the sample corresponding to that used in estimating model 3 (results not reported). The LR test statistic of 2066.20 was computed with a p-value of less than 0.0001, justifying the added complexity.

In summary, under model 3 (as reported in Table 26):

1) Arrival delay coefficients are positive and significant for all journey purposes and in both seat and no-seat scenarios, though typically larger for non-seated passengers (with not statistically significant difference for business travellers).
2) The impact of scheduled journey length is generally negative (though insignificant for commuters and business travellers to London).
3) Discerning the complementary effect of having a seat on how journey length affects delay perception is more challenging, given that the probability of perceiving a delay increases in a different way for seated and non-seated passengers. Hence, a reduction in the probability of perceiving a delay due to increasing journey length for seated and standing passengers is from a different level.
4) The impact of departure delay is significant and positive, meaning that for a given length of arrival delay, the probability of perceiving a delay is larger if the train also departs late.

Table 26 Estimated coefficients from the binary logit model of delay perception (model 3 and model 3a with all levels of interacted variables)

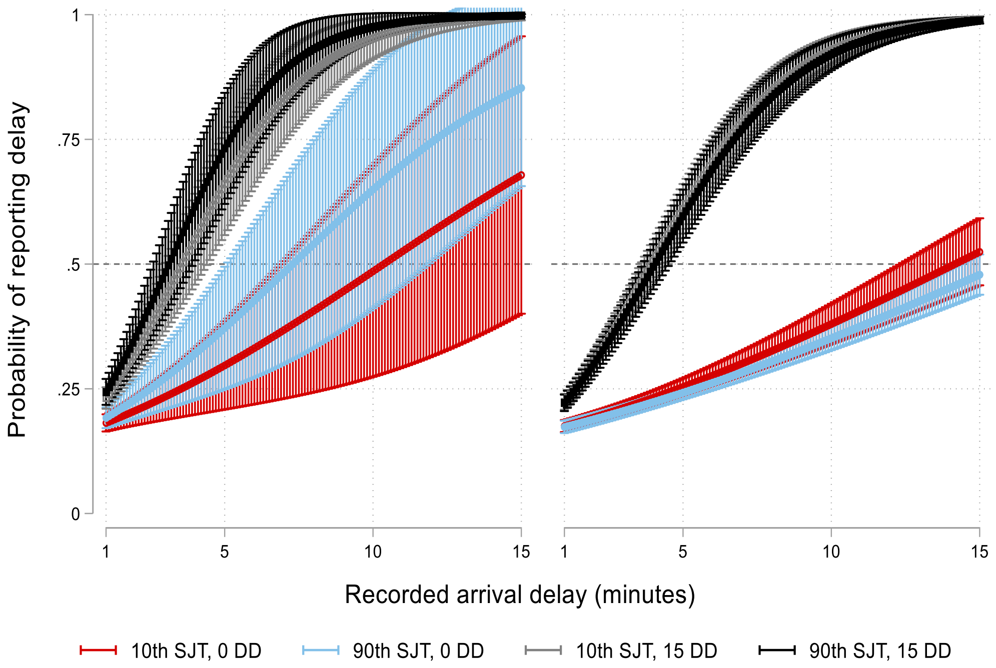
	(3)	(3a)
Constant	-1.666***	-1.773***
	(-34.68)	(-4.09)
BnL	0.338***	0.336
	(4.60)	(0.65)
CL	0.783***	0.433
	(12.71)	(0.95)
CnL	0.966***	0.843
	(13.91)	(1.80)
LF	0.00961	0.213
	(0.14)	(0.44)
LR	0.0416	0.312
	(0.75)	(0.67)
Seat=1		
BL		-0.488
		(-1.11)
BnL		-0.528
		(-1.77)
CL		-0.462**
		(-2.75)
CnL		-0.416*
		(-2.15)
LF		-0.810***
		(-3.72)
LR		-0.730***
		(-4.00)
Arrival delay (Seat=0)		
BL	0.150**	0.0941
	(2.76)	(1.41)
BnL	0.192***	0.110**
	(4.99)	(2.78)
CL	0.229***	0.140***
	(8.55)	(4.60)
CnL	0.336***	0.132***
	(10.32)	(3.48)
LF	0.290***	0.125***
	(9.84)	(3.41)
LR	0.247***	0.126***
	(10.83)	(5.38)
Arrival delay (Seat=1)		
BL	0.120***	0.161***
	(9.94)	(11.21)
BnL	0.145***	0.152***
	(9.16)	(9.03)
CL	0.115***	0.159***
	(7.61)	(9.25)
CnL	0.189***	0.116***
	(8.83)	(5.74)
LF	0.152***	0.151***
	(11.96)	(11.83)
LR	0.144***	0.159***
	(19.86)	(20.39)
Arrival delay x SJT (Seat=0)		
BL	0.000512	0.00112
	(0.86)	(1.24)

BnL	-0.000738*	0.0000925
	(-2.33)	(0.22)
CL	-0.0000537	0.000278
	(-0.07)	(0.31)
CnL	-0.00392***	-0.00174
	(-4.37)	(-1.47)
LF	-0.000933**	0.000474
	(-3.15)	(0.88)
LR	-0.000727***	-0.000102
	(-4.05)	(-0.51)
Arrival delay x SJT (Seat=1)		
BL	-0.0000924	-0.0000512
	(-1.22)	(-0.46)
BnL	-0.000489***	-0.0000159
	(-5.19)	(-0.12)
CL	-0.000152	-0.000223
	(-0.61)	(-0.66)
CnL	-0.00168***	0.000307
	(-4.80)	(0.70)
LF	-0.000482***	0.0000929
	(-5.51)	(0.77)
LR	-0.000298***	0.0000536
	(-7.55)	(0.99)
SJT (Seat=0)		
BL		-0.00129
		(-0.21)
BnL		-0.00435
		(-0.95)
CL		0.0129**
		(2.71)
CnL		0.00851
		(1.31)
LF		-0.00601
		(-1.45)
LR		-0.00142
		(-0.62)
SJT (Seat=1)		
BL		0.0000570
		(0.06)
BnL		-0.00302**
		(-2.83)
CL		0.00627**
		(2.94)
CnL		-0.00507
		(-1.68)
LF		-0.00284**
		(-2.62)
LR		-0.00255***
		(-5.30)
Departure delay		
BL		0.465***
		(23.27)
BnL		0.461***
		(22.28)
CL		0.454***
		(27.05)
CnL		0.450***
		(24.11)

LF		0.466***
		(28.99)
LR		0.453***
		(42.84)
Departure delay x arrival delay		
BL	0.0204***	-0.0170***
	(13.27)	(-14.23)
BnL	0.0153***	-0.0166***
	(9.62)	(-14.85)
CL	0.0119***	-0.0186***
	(8.64)	(-19.98)
CnL	0.00792***	-0.0167***
	(4.94)	(-17.55)
LF	0.0106***	-0.0183***
	(9.56)	(-21.14)
LR	0.0147***	-0.0167***
	(19.54)	(-29.23)
N	48793	48793
LL	-26973.4	-24095.8
Pseudo R^2	0.167	0.256

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced; SJT: scheduled journey time, Seat=0 represents a standing passenger

The predicted probabilities are shown below in Figure 36-Figure 38 for each of the journey purposes under the scenarios where passengers did or did not have a seat, for the 10th and 90th percentile of journey length and the departure delay of respectively 0 and 15 minutes. The complexity introduced by the multiple interactions between numerous variables makes the direct comparisons complicated. However, it can be seen that generally, the predicted probabilities of perceiving a delay are always highest for a given length of delay at arrival if the passenger was standing, the train departed late and the journey was relatively short. This is also demonstrated by the estimated delay length thresholds at the p=0.5 in Figure 39 below. Looking at the shorter journeys with a large departure delay and in the case when the passenger was standing, the estimated thresholds are between 2.0 minutes for non-London commuters to 3.9 minutes for leisure travellers on full fare. On the other hand, the estimated thresholds are the largest for the longer journeys with no delay at the departure and when the passenger was able to find a seat, between 6.4 minutes for non-London commuters to 19.9 minutes for non-London business travellers. This would be indicative of the impact that both journey quality, length and delay at departure can have on the perception of arrival delay, also suggesting that the impact of these travel attributes is much smaller on the commuters.



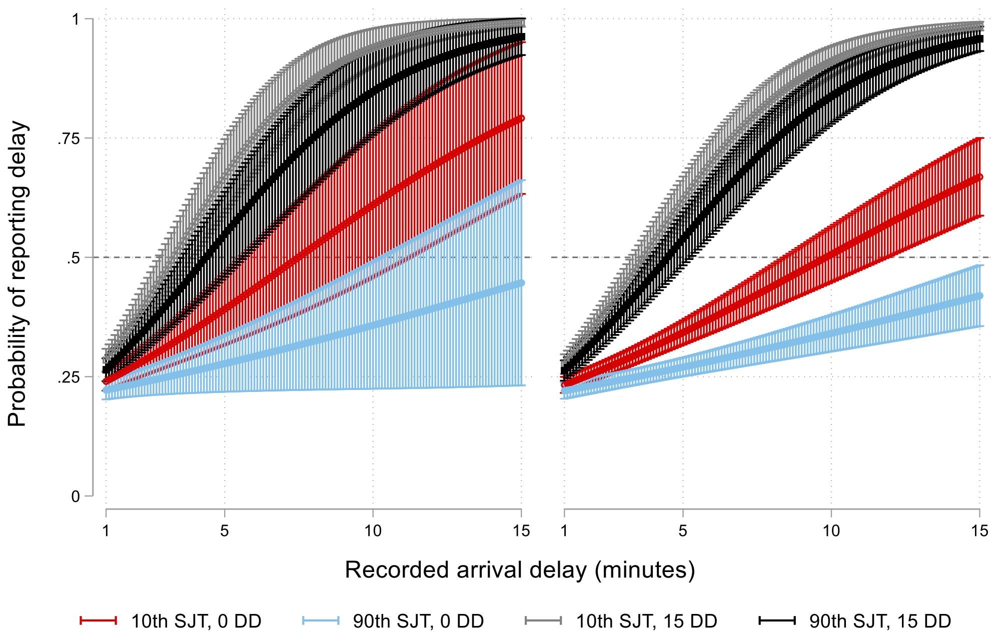


Figure 36 Estimated probabilities of delay perception for model 3 for business London (top) and non-London (bottom) under the ‘no seat’ (left) and ‘seat’ scenario (right)

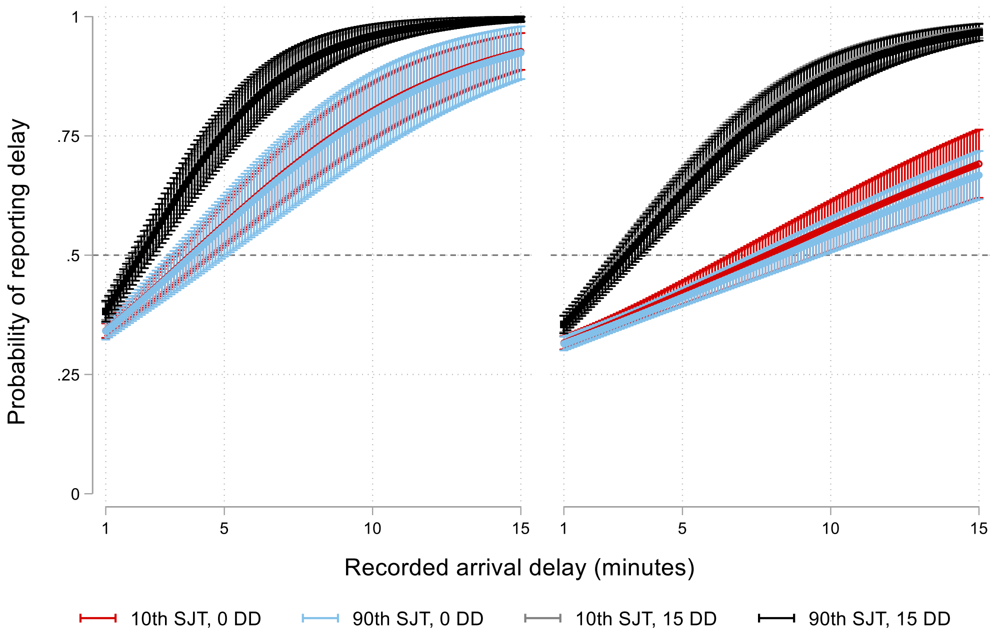


Figure 37 Estimated probabilities of delay perception for model 3 for commute London (top) and non-London (bottom) under the 'no seat' (left) and 'seat' scenario (right)

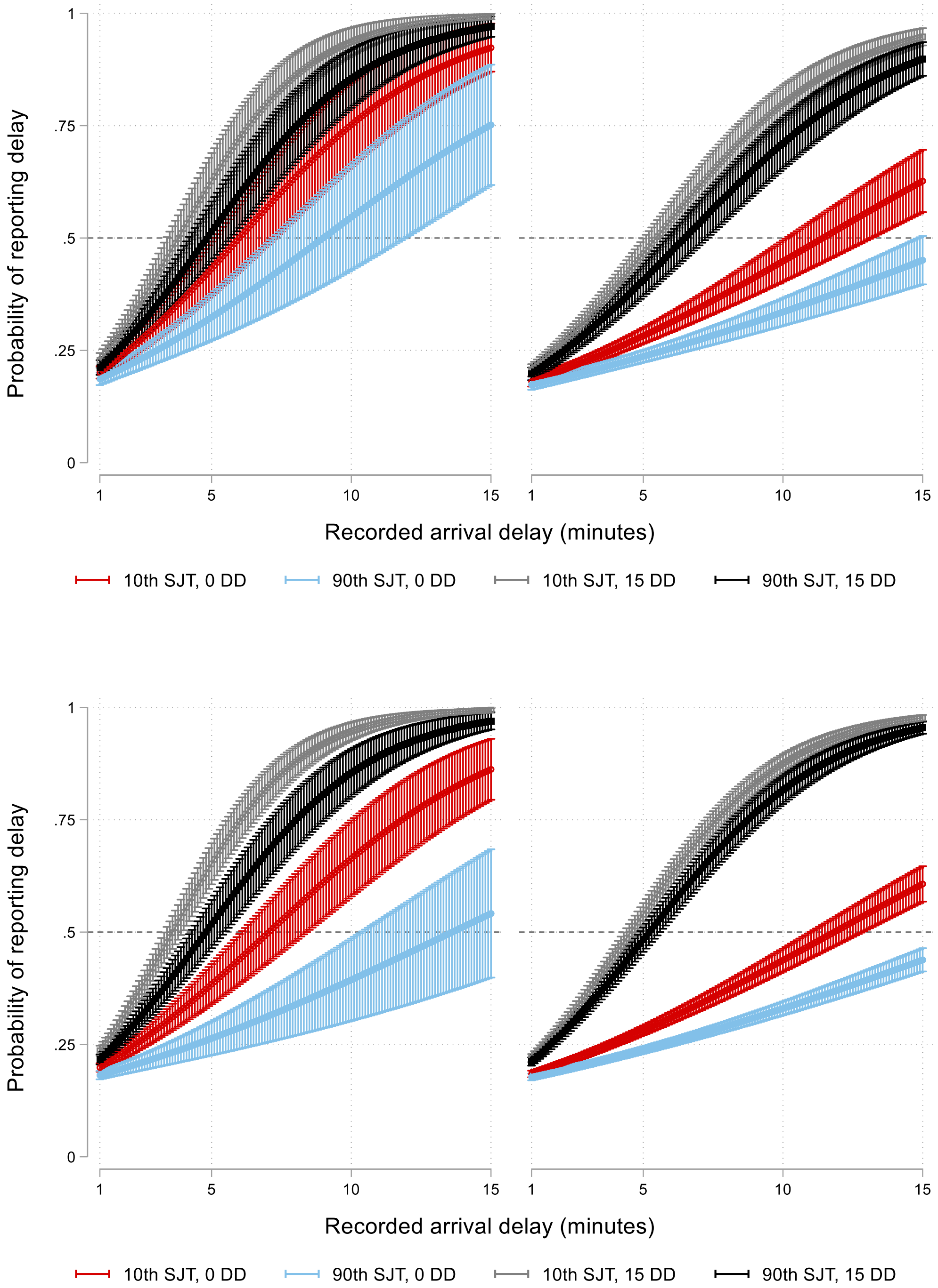


Figure 38 Estimated probabilities of delay perception for model 3 for leisure full (top) and reduced (bottom) under the 'no seat' (left) and 'seat' scenario (right)

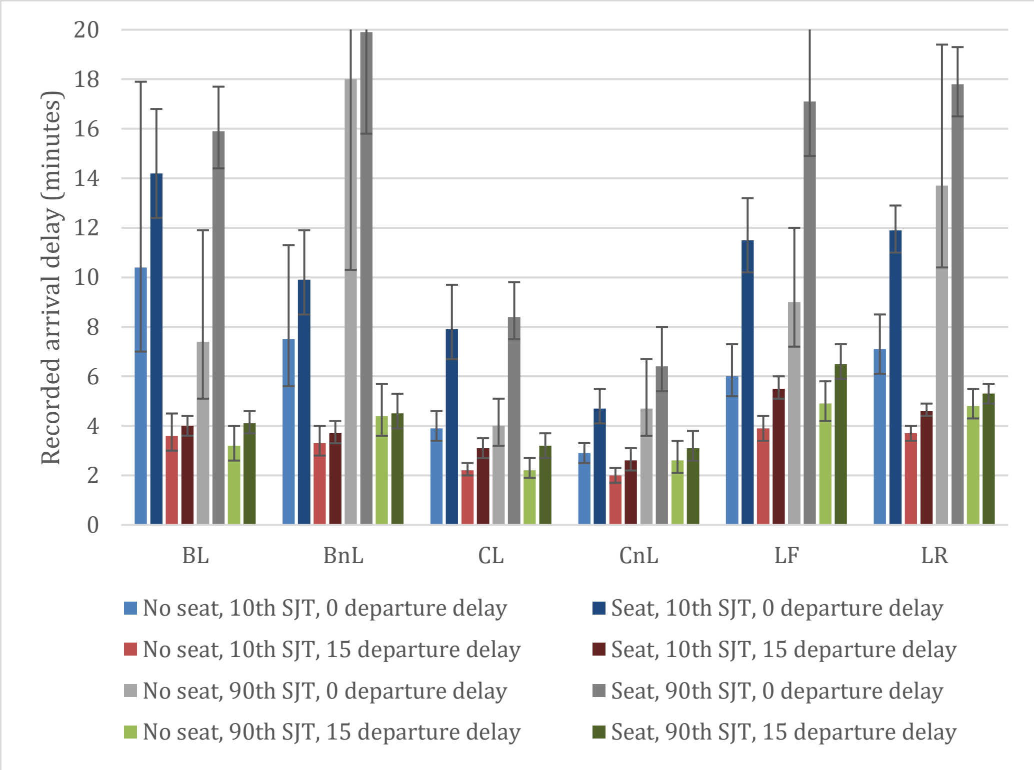


Figure 39 Estimated delay length thresholds for p=0.5 for different journey quality, SJT, and departure delay scenarios (based on model 3)

Considering the complexity of the modelling approach and the number of continuous variables and interactions between them, the summary below aims to highlight the key conclusions regarding the impact of each of the variables on the delay perception of different types of travellers.

Business and leisure

The probability of perceiving an arrival delay by business and leisure travellers typically largely increases with the length of departure delay. As the departure delay increases, the equivalent delay at arrival required to reach 0.5 probability of delay perception decreases by up to 15 minutes, depending on journey length and seat availability. This also suggests that all these attributes have a large and complementary effect on the delay perception. Travellers on longer journeys may incorporate larger safety buffers to increase the probability of arriving to the destination before their scheduled activity (i.e. meeting or leisure activity) starts and, thus are not as sensitive to the smaller delays. Moreover, seated travellers are also more likely to use their travel time productively (i.e. on work or leisure activities) which can reduce their focus on on-time versus late arrival.

Commute

Commuters' perception of arrival delay is much less likely to be impacted by factors such as journey quality or length. However, the probability of perceiving a delay generally decreases if a traveller is seated and for longer journeys, and increases with length of delay at departure. Nevertheless, these impacts are typically less pronounced for commuters. Comparing with other travellers, for a given length of delay (all else being equal), commuters' probability of perceiving a delay is typically larger than for the other passengers, but commuter journeys are usually also relatively short. To reach a 0.5 probability of perceiving a delay, typically an arrival delay of 2-8 minutes is needed. This threshold decreases with increasing length of delay at departure. Having a seat has a significant and negative impact on the delay perception of London commuters. Non-London commuters' perception of delays is less affected by journey quality and departure delay. This could be a result of the differences in the safety buffers included in their journeys or differences in service offering that may be characteristic for non-London services (i.e. lower frequency of departures, meaning that any possible disruption has a larger impact on their journeys).

It is highlighted that the comparison between the proposed model (model 3) and the fully specified model (model 3a) that includes all levels of interacted variables is complex due to the large number of interactions included. In the case of model 3a, some of the estimated coefficients do not have the expected sign (e.g. the coefficient on the interaction between arrival and departure delay is negative) or signs differ between journey purposes (e.g. the coefficients on scheduled journey lengths). While it is expected that the magnitude of the impact may differ between passengers, the direction of the relationship is expected to be uniform across all journey purposes and generally in line with the relationships estimated by model 3. The complication of this is clearly visible in Annex II in Figure 68. When looking at the plotted estimated probabilities of delay perception, it is clear that model 3a (estimated with inclusion of all levels of the interacted variables) fails to correctly predict the direction of the studied relationship. The estimated probability of perceiving a delay is suggested to decrease with increasing levels of delay at arrival at larger values of delay at departure. Moreover, in this case, the average marginal effects also suggest that the impact of a minute of delay at departure is larger than the impact of a minute of delay at arrival. This is expected to be due to the presence of multicollinearity in the model with all levels of interacted variables included. The variance inflation factor (Marquardt, 1970) for model 3a was calculated to be 46.6 as compared to 3.9 for model 3 in its original form, suggesting that the model with all levels of the estimated variables suffers from high level of multicollinearity. Given this, the more parsimonious model (i.e. model 3) which also yields interpretable and plausible results remains the preferred model.

Model performance

The estimated models have McFadden's R^2 (McFadden, 1974) of around 0.1-0.2 what is of a similar magnitude to the values reported for the ordered logit models of journey satisfaction in Monsuur et al. (2021) which also used NRPS data. It is clear that while a meaningful relationship between the variables is found in the data and statistically significant estimates can be computed, the model overall is not able to capture a very large proportion of the variance in the data.

To investigate model performance, the estimated models were used to predict the outcomes (i.e. reports of delay perception) for the respondents in the samples. As summarised in Table 27, all the models correctly predict over 70% of responses with model 3 performing best (correct prediction rate of 73.41%). While determining whether model performance is satisfactory may depend on the context of the study, in this case, for almost 3 in 4 travellers, the model correctly predicts their ability to perceive a delay. There are two additional measures that are typically reported when commenting on model performance (e.g. Harris, 2021):

1) Sensitivity, defined as the ratio of correctly predicted positive outcomes (delay reports) to all positive outcomes. This ratio is also highest for model 3, at 77.77%.

2) Specificity, defined as the ratio of correctly predicted negative outcomes (on-time arrival reports) to all negative outcomes. This ratio is highest for model 2, at 72.72% with minimal differences between the models.

Additionally, it was investigated what the minimal predicted perceived and maximum predicted unperceived delay lengths are for each of the estimated models. In the dataset, there are both cases where a minimum delay length of 1 minute is perceived and the maximum delay length of 30 minutes is unperceived. In the case of the estimated models, it is noted that some heterogeneity in the perception of delays is expected that needs to be controlled for, as is done by the inclusion of additional variables in models 2 and 3. Model 1 is less capable of predicting the heterogeneities in the delay perception reports. Models 2 and 3 are able to better capture the additional effects of journey length and quality as well as delay at departure what, in many cases, explains why smaller delays are perceived or larger delays remain unperceived. However, there remains a sizeable proportion of perceived delays that were predicted to be unperceived. In the dataset, around 35-37% of respondents reported a delay whereas the proportion of predicted perceived delays is at 18-

20%. Sensitivity analysis (Annex I) tries to investigate whether the reasons for some wrong predictions may be attributed to data inaccuracies or heterogeneity in travellers.

Table 27 Model performance

Delay reported	Predicted	Model 1	Model 2	Model 3
Yes	**Yes**	12.95%	13.93%	15.81%
Yes	**No**	22.93%	21.95%	22.08%
No	**Yes**	5.39%	5.61%	4.52%
No	**No**	58.73%	58.51%	57.6%
Total correct predictions		71.68%	72.44%	73.41%
Total reported late (data)		35.88%	35.88%	37.89%
Total predicted reporting late		18.34%	19.54%	20.33%
Sensitivity (true positives to all positives)		70.61%	71.29%	77.77%
Specificity (true negatives to all negatives)		71.92%	72.72%	72.29%
Minimum delay predicted perceived		5	3	2
Maximum delay predicted unperceived		11	30	30

6.4.4. Summary

This section introduced the concept of delay perception by examining how travellers' probability of perceiving a delay changes with:

- increasing delays at arrival and departure,
- passengers being seated or standing,
- journey length and
- journey purpose.

In line with the expectations, the following effects have been suggested:

- The probability of perceiving a delay increases with increasing length of delay at arrival.
- For any given length of delay at arrival, the probability of perceiving it also increases with increasing delay at departure.
- For any given scenario, it can be expected that travellers are more likely to perceive a delay if they are standing (e.g. travel in crowded conditions).
- Typically, the length of the journey contributes to a lower chance of perceiving a delay.

- Journey type was indicated to have a strong effect on delay perception as typically, commuters' probability of perceiving the same level of delay is larger.

Whilst some of the estimated results are, in fact, intuitive, studying delay perception may provide an explanation for differing impacts of delays on different types of travellers as well as explain the possible non-linearities in the delay impacts, which can be a result of the suggested relationships. Before proceeding to discuss these as part of the remaining chapters, the following sections will introduce the additional approaches to studying delay perception, by looking at:

- passengers reporting being delayed when no delay was matched and
- differences in the reported and recorded delay lengths

6.5. Delay perception when no delay is recorded

The previous section investigated how the probability of perceiving a delay changes with increasing levels of recorded delays. However, as described in section 5.2.2, in 5.8% of responses, a traveller reported arriving late, but no delay was recorded at the destination station.

The main limitation of the NRPS dataset is the inability to define the exact journey that passenger was planning and expecting to make as compared to their actual experience. This means that while the information provided by passengers regarding their origin, destination and services used is useful as it allows matching passenger to a specific service they travelled on, it is impossible to verify whether the actual journey was the same as the planned one. Moreover, in the case of cancellations and multi-leg journeys, there is a possibility that passenger responses reflect on the whole journey rather than a specific journey leg, which can lead to some reporting inaccuracies.

Due to these reasons, the responses where despite on-time performance being matched, a passenger reported late arrival, were removed from the dataset as noted in section 5.5. However, some additional analysis may be conducted to understand whether these reports could have been impacted by the delay at departure that was possibly later recovered (as no delay was recorded at arrival). Delay at departure may affect judgment and, therefore, there is a possibility that passengers arriving to their destination on time may perceive journeys as delayed due to late departure.

On average, a departure delay of 0.38 minutes (95% confidence interval range of 0.376-0.390) was registered for the responses where passengers reported on-time arrival as compared to 1.46 minutes (95% confidence interval range of 1.412-1.501) for the responses where passengers reported late arrival. The difference between the reported means is statistically significant (p=0.000). Moreover, Table 28 summarises the distribution of recorded lengths of delay at departure for responses where no delay at arrival was matched and passenger either reported arriving on-time or late. In the former case, over 3 in 4 respondents who reported on-time arrival did not experience a delay at neither departure nor arrival. In the latter case, when passengers reported late arrival, but no arrival delay was matched, 1 in 2 respondents were matched a delay at departure with almost 15% of responses being characterised by a departure delay of over 4 minutes. This provides some indication that departure delay may also affect delay perception in the cases where the train arrives to the destination as scheduled. While the difference in the average departure delay may provide some explanation for this phenomenon, it would indicate that the thresholds of departure delay perception are smaller than in the case of the previously modelled delay perception for the responses affected by arrival delays. A simple logit model was run to investigate this further. This model imitates model 1 from Table 25, but the outcome variable now refers to delay reports in the case of no recorded arrival delay. Moreover, the delay perception is explained by the length of delay at departure, rather than at arrival (results presented in Table 29).

In all the cases, the departure delay coefficients are positive and significant, indicating on the departure delay affecting travellers' judgment and increasing the likelihood of perceiving a delay even if the train arrives to the destination as scheduled. The model predicts the outcome correctly in almost 80% of responses, though it better predicts the outcome for passengers who did not report a delay in this case (95.3% correct) as compared to passengers who did report a delay (21.7% correct). This might indicate on the existence of some additional reasons (likely data errors) why passengers who did not experience a delay at arrival reported otherwise as departure delay may be one of the explanations, but likely not the only one.

Table 28 Distribution of recorded departure delays for passengers who reported on-time or late arrival

Recorded delay length at departure	Proportion of passengers who reported on-time arrival when no arrival delay was matched	Proportion of passengers who reported late arrival when no arrival delay was matched
0	76.84%	48.17%
1-3	21.69%	37.53%
4-6	1.30%	11.08%
>6	0.17%	3.22%

Table 29 Logit estimates of delay perception where delay is recorded at departure, but not at arrival

	(1)
Constant	-2.657***
	(-23.98)
BnL	0.173
	(1.05)
CL	0.898***
	(6.71)
CnL	1.074***
	(7.21)
LF	0.0853
	(0.61)
LR	-0.123
	(-0.97)
Departure delay	
BL	0.587***
	(14.83)
BnL	0.505***
	(12.32)
CL	0.608***
	(16.47)
CnL	0.506***
	(9.89)
LF	0.439***
	(14.39)
LR	0.528***
	(25.89)
N	19130
Log-likelihood	-8880.8
Pseudo R^2	0.126
% correct	79.76%

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced

Having investigated travellers' ability to perceive a delay, the interest now lies in understanding how accurately the respondents reported the lengths of delays they experienced which will be investigated in the following section.

6.6. (Mis)perception of delay lengths

Previous sections were concerned with explaining the reasons why:

1) delays are/are not perceived in the cases when delay is recorded (section 6.4) or
2) why they might be perceived despite no delay being recorded (section 6.5).

It may be expected that travellers can not only misperceive the fact of delay occurring but also its length. The focus of the analysis is on the subset of the dataset with travellers who reported late running and were also matched a delay and only the first three survey waves where passengers needed to state experienced delay length in minutes, instead of choosing the delay length categories (as discussed in section 5.3). The analysis was conducted for responses where the recorded arrival delay was up to 30 minutes and passengers reported that they were delayed by up to 60 minutes, only for passengers who reported not changing trains (to improve accuracy). The methodology is based on analysing the distribution and summary statistics for the reported and recorded delay lengths with no additional econometric models introduced.

Generally, there could be two possible reasons for the existence of differences between reported and recorded lengths of delays, related to delay perception or rounding errors. There is a large body of research investigating rounding of survey responses (e.g. Manski and Molinari, 2010; Giustinelli et al., 2019). For example, in the case of responses reported on a 1-100 scale, values that are not multiples of 5 occur relatively infrequently (Giustinelli et al., 2019). In transport, as noted by Rietveld (2001), Sato and Maruyama (2020) and Sanko and Iriguchi (2022), rounding errors are often present in travel surveys as respondents round their departure and arrival times to the nearest multiples of 5, 10, 15, 30 and 60 minutes. As noted by Rietveld (2001), the departure times are more likely to be rounded than arrival times, what can be a result of a larger penalty for late arrival due to fixed schedules. In the case of the national transport survey from the Netherlands (Rietveld, 2002), only up to 15% of reported times were not multiples of 5, suggesting that rounding is a very common phenomenon in travel surveys. As suggested by Rietveld (2001), this highlights the potential for biases and errors in using reported travel time data from surveys.

Table 30 and Table 31 below provide a summary of respectively average reported delay length for every possible value of recorded delay and average recorded delay length for every possible value of reported delay with the distribution of reported delay lengths presented in a histogram in Figure 40.

Table 30 Summary of reported delay lengths for each of the recorded delay lengths

	Reported delay length (minutes)					
Recorded delay	**Mean**	**Max**	**Min**	**SD**	**Median**	**N**
1	8.87	60	1	9.14	5	745
2	7.84	60	1	8.50	5	710
3	8.63	60	1	9.48	5	647
4	8.26	60	1	7.69	5	628
5	8.27	60	1	8.14	6	557
6	9.13	60	1	7.36	7	399
7	9.86	58	1	7.31	8	366
8	10.05	45	2	6.75	9.5	346
9	10.49	60	1	7.68	10	276
10	12.89	60	2	9.50	10	213
11	12.25	60	1	7.42	10	181
12	12.78	60	1	7.37	10	175
13	13.40	50	1	6.68	12	129
14	14.97	50	1	7.56	15	134
15	14.21	30	1	5.59	15	107

Table 31 Summary of recorded delay lengths for each of the reported delay lengths

	Recorded delay length (minutes)					
Reported delay	**Mean**	**Max**	**Min**	**SD**	**Median**	**N**
1	6.22	30	1	7.01	3	79
2	2.91	29	1	2.77	2	354
3	3.87	30	1	3.74	3	489
4	4.60	30	1	3.85	4	343
5	4.73	24	1	3.49	4	1165
6	5.97	22	1	3.58	5	327
7	6.13	29	1	4.02	5	252
8	7.00	28	1	4.14	7	227
9	7.27	20	1	4.18	7	75
10	7.46	29	1	4.94	7	1166
11	8.88	22	1	4.79	9.5	56
12	9.86	22	1	5.41	10	109
13	11.59	29	1	5.39	12	49
14	12.52	28	1	5.18	13	31
15	10.44	29	1	6.55	10	543

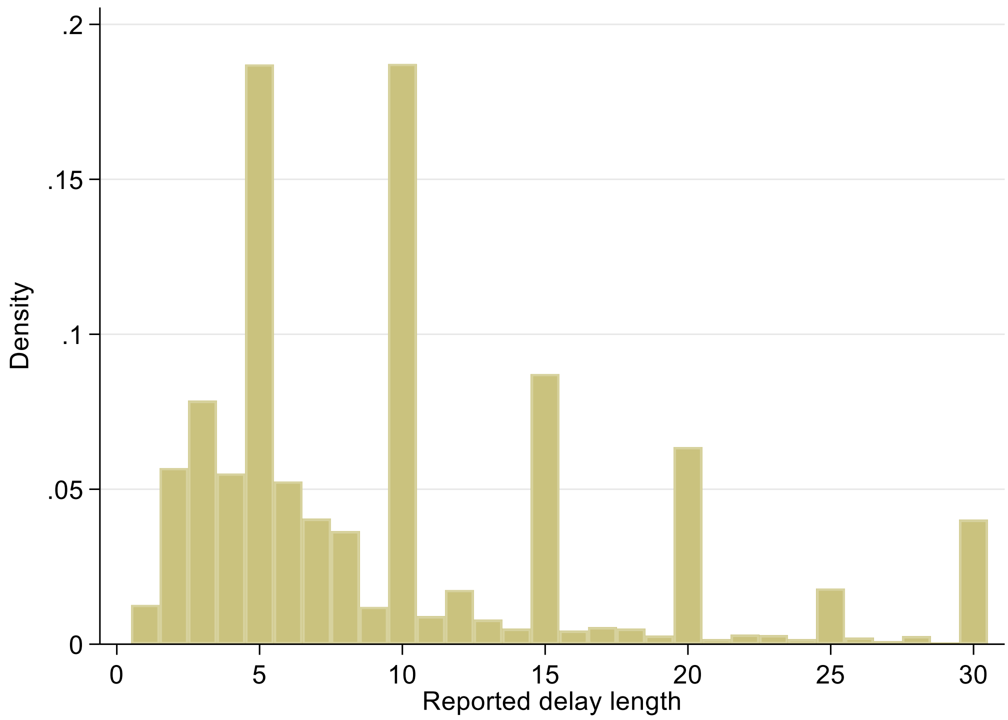


Figure 40 Distribution of reported delay lengths

In line with expectations, passenger-reported lengths of delays are scattered around the closest 5 or 10 minutes. In the case of recorded delays, the lengths that are multiples of 5 represent 15.8% of all responses while 58.3% of passengers reported their delay lengths to be a multiple of 5. In the case of Rietveld (2001), the rounding to the nearest multiple of 5 was more frequent, characterising up to 85% of travel time reports. However, there are some key conceptual differences between departure or arrival time (reported in Rietveld, 2001) and the delay length (analysed here). Moreover, it seems that there are relatively few delay length reports concentrated around 2-8 minutes. Nevertheless, as noted in section 5.2.2, the distribution of recorded delays is skewed towards smaller delays as these are very frequent. In the case of the studied subsample, 11.6% of recorded delays are of only 1 minute, 32.7% are within 3 minutes and 68.5% are within 8 minutes. In the case of reported delays, these proportions are respectively 1.3%, 14.8% and 51.2%. A relatively large number of delay reports (over 30%) in the case of smaller delays is not a multiple of 5. However, this is still lower than the proportion of the smaller recorded delays that are not concentrated within multiples of 5 (almost 60% of all delays).

The plots below show the average lengths of reported delays for each of the recorded delay length categories in Figure 41 and the average lengths of recorded delays for each of the reported delay length categories in Figure 42. This suggests that the average length of delay is over-reported for recorded delays of up to 15 minutes. However, the difference generally becomes smaller with increasing recorded length of delay. In the case of reported delay lengths, the average recorded delay length (as well as the median as shown in Table 31) is very close to the reported delay lengths for delays of between 2 to 8 minutes. This would indicate that passengers who perceive such delays typically also perceive their lengths more accurately. The increased accuracy may be resulting from the fact that smaller delays, while generally less likely to be noticed, may be more likely to be perceived by passengers very sensitive to late running, especially commuters. Commuters are also likely to be more familiar with scheduled departure and arrival times of services that they regularly use what can, in turn, enable them to more quickly notice any deviations from the timetable and estimate any delays more accurately. It is also evident that in the case of reported delays of 1 minute, the average recorded delay is above 6 minutes, which is due to data errors or more frequent rounding of delays down to 1 minute in the case of some smaller delays. In the case of reported delays between 8 and 14 minutes, the average recorded delay is typically very close to the reported delay length, however, slightly lower. In the case of larger reported delays, especially above 20 minutes, the relationship becomes more difficult to follow as the number of responses largely decreases.

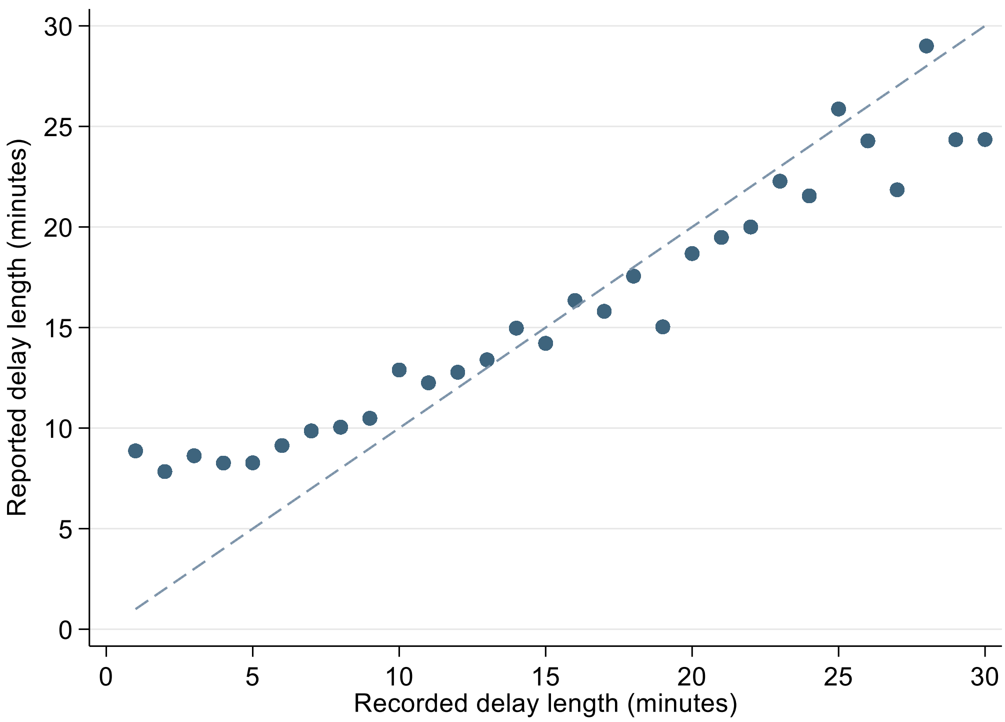


Figure 41 Average reported delay length for each of the recorded delay length categories

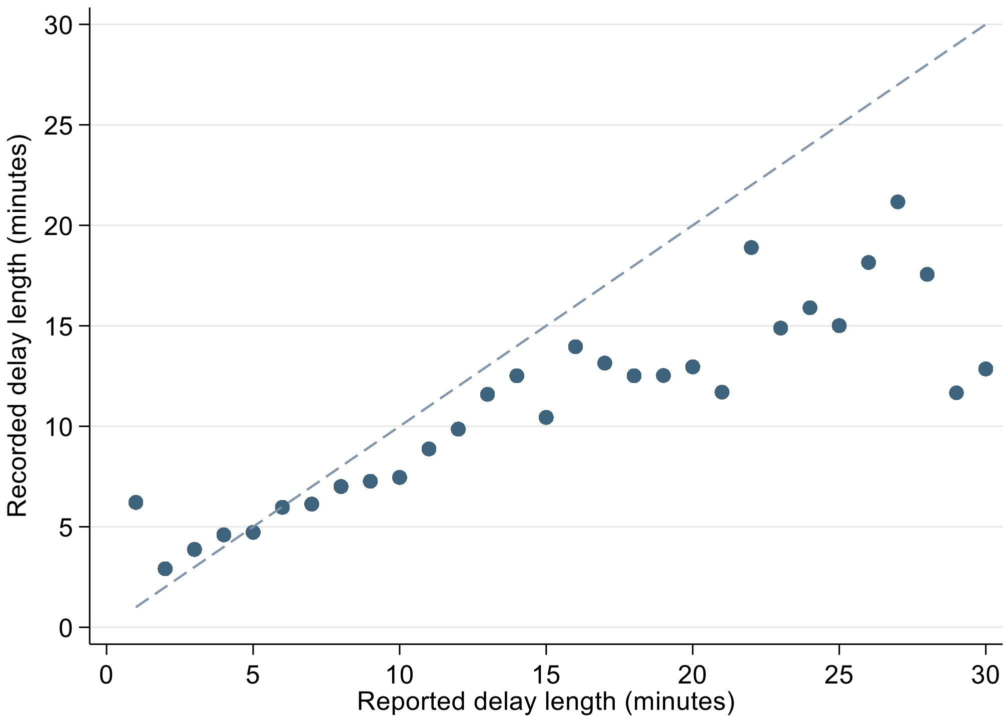


Figure 42 Average recorded delay length for each of the reported delay length categories

Upon suggestions that travellers who notice smaller delays may be more accurately perceiving their duration, additional investigation was conducted to understand the possible differences between different types of travellers. As suggested by the modelling in section 6.4, commuters are more likely to perceive smaller delays. Therefore, Figure 43 looks at the average differences between reported and recorded delay lengths for commuters and other travellers. It can be seen that there does not seem to be a systematic difference in the accuracy of delay lengths as reported by commuters and other travellers. The only noticeable difference is the more accurate reporting of the delays in the case of the reports of 1-minute delays. Hence, this suggests that the improved accuracy of reported delay lengths for smaller delays may be a result of increased focus on performance by the travellers with increased sensitivity to lateness (not only commuters).

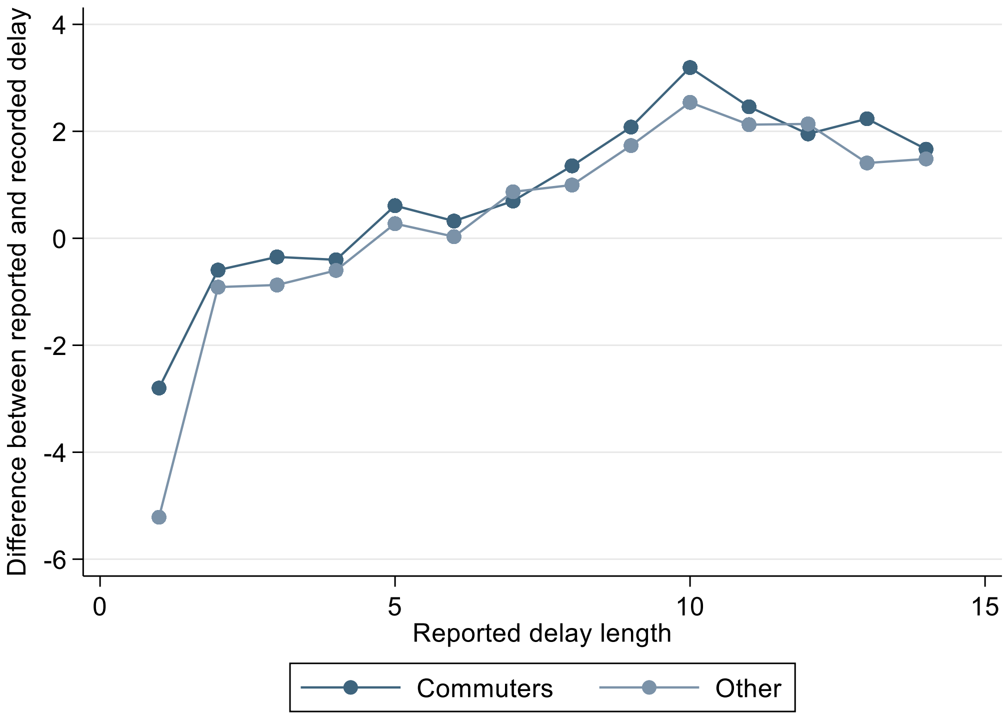


Figure 43 Average difference between reported and recorded delay length for commuters and other travellers

6.7.Conclusions

Motivated by the suggestions by Wardman and Batley (2022) and Rong et al. (2022), highlighting the importance of understanding how delays are perceived by passengers, this chapter investigated how the levels of recorded delays affect passengers' delay reports.

Passenger delay reports from the National Rail Passenger Survey in the United Kingdom were matched to an operational dataset to relate passengers' perception of late running to actual performance. Passengers' ability to perceive a delay was first modelled as a binary outcome, explained by the recorded length of delay registered at arrival and departure points, while also controlling for journey purpose, scheduled journey time, and journey quality. The results indicate that journey quality, length and delay at departure generally all have an impact on the perception of final performance with the probability of perceiving a delay generally suggested to increase with delay at arrival and departure but decrease with journey time and better journey quality. However, the effect of journey quality and length is generally larger if a passenger faces smaller delays at departure.

Commuters are suggested to be the most sensitive travellers with respect to delays, typically being able to perceive arrival delays as small as 2 to 8.5 minutes. For business travellers, the respective thresholds are between 3.2 to 19.9 minutes and for leisure travellers 3.7 to 17.8 minutes. In fact, the estimated thresholds are very similar for the two latter groups of travellers. Commuters may include shorter buffer times, their journeys are typically shorter and are also expected to be more familiar with the timetables what can result in their better ability to notice delays earlier than other travellers.

The analysis highlights the large impact of departure delay on delay perception as it seems that passengers waiting for a delayed train at the origin station may be more likely to think that their journey was delayed what can be a result of the increased uncertainty and discomfort. Similarly, passengers who departed on time and were able to find a seat could use their travel time more productively, in turn, reducing their ability to perceive a delay (i.e. a longer delay is needed to achieve the same level of delay perception) and especially so for the longer journeys. However, the impact of departure delay on delay perception and satisfaction may, in fact, be different and future studies might want to incorporate departure delay into modelling the impacts of delays on passenger satisfaction.

Moreover, an attempt was made to investigate the reports of late arrival in the case where no arrival delay was recorded. This was done to better understand whether such cases are due to:

1) data errors due to differences between planned, experienced, recorded and reported journeys or
2) perception of delay being strongly impacted by the delay at departure.

The analysis suggested that the delays that are perceived in the case of recorded on-time arrival are possibly a result of both judgment impacted by recorded delay at departure and data errors. On one hand, the probability of perceiving a delay in the case of no recorded arrival delay was suggested to be larger for cases where a delay at departure was matched. However, still almost 1 in 2 passengers who reported late arrival and were not matched a delay at arrival, were also not matched a delay at departure. Further investigation could possibly explore whether these erroneous reports may have been a result of stopping patterns (as discussed in Rong et al., 2022), the inclusion of recovery times (as discussed in Ojeda-Cabral et al., 2021) or delays that happened while being on-board that were fully recovered.

The next line of investigation was the analysis of perception of delay lengths for travellers who reported being late and were also matched a delay. The investigation generally indicated that similarly to reports of departure or arrival times in travel surveys, passengers tend to round the lengths of delays to the nearest multiple of 5 minutes. However, less commonly than when reporting departure and arrival times, likely due to a larger penalty related to late arrival, as suggested by Rietveld (2001). Travellers generally quite accurately predict delay lengths between 2-8 minutes whereas longer delays are more often overstated. This may be a result of correlation between the ability to perceive a smaller delay and higher accuracy of reporting delay lengths for passengers who are more sensitive to delays due to their increased focus on train performance. However, the analysis generally highlighted the previously reported implications related to only using passenger reports in the analysis of travel patterns as some discrepancies were found that result from either misperceptions or data errors.

Future research directions and policy implications

The research conducted as part of this chapter fills the important gap in the literature by improving the understanding of delay perception. The predicted probabilities of delay perception and the respective thresholds suggest that very short delays are unlikely to be perceived. These thresholds could perhaps be used in formulating performance targets or further research into the non-linearity of delay impacts as, in line with previous suggestions, for a delay to have an impact on passengers, it would typically need to be perceived first. However, it is important to distinguish between delay perception and delay impacts on utility (or passenger satisfaction). The impact of delays on satisfaction was previously studied using the NRPS dataset by Monsuur et al. (2021) where the delay threshold of 30 minutes was suggested to be the cut-off point after which passengers are unlikely to remain satisfied with their journeys. That piece of work, however, only focused on passengers who were able to perceive a delay. This thesis, on the other hand, aims to test the hypothesis that passengers are not always able to notice the shorter delays which would suggest that the impact of unperceived delays is perhaps smaller. Therefore, the most obvious approach is to look at the relationship between delay lengths and reported satisfaction. This will be done as part of the following chapter and subsequently (in section 7.5) the delay perception thresholds will be contrasted with the satisfaction thresholds what can help understand the differences between perceived delays and such that have a detrimental impact on passenger satisfaction.

While this research used data on delay perception matched to an operational dataset, it is believed that the dataset would benefit from a more detailed description of planned and actual journeys, including the services that a passenger was planning to travel on and actually travelled on, and any possible interchanges. In addition to such information, data on fares, headways and average performance could be a useful addition, allowing for a more detailed investigation of the studied relationships. The dataset used as part of this analysis may be prone to some discrepancies resulting from the imperfect information about passenger journeys.

Unfortunately, it was impossible to study the impact that on-board and at-station notifications or access to real-time delay information may have on passengers' ability to perceive delays, as it is difficult to understand whether, how and when a passenger was informed about any disruption. However, it is believed that the provision of information about late running may also be key in determining passengers' perception of performance.

This work has increased understanding of passengers' perception of delay, but it is worth noting that the policymakers' and rail operators' objective should not, therefore, target travellers' perception of delays as such approach is unlikely to be welfare maximising. The major benefits from the perception research would be in incorporating delay perception into research focusing on the impacts of lateness on passenger satisfaction and ultimately demand. This could focus on understanding the possible inherent non-linearities in delay impacts where the delay perception may be one of the explanations for such relationships being present, possibly leading to a discussion about how smaller versus larger delays should be treated in designing performance metrics, compensation schemes and appraisal of schemes looking at improving reliability.

Chapter 7
Impacts of delay on travellers' satisfaction

7.1. Introduction

The impact of delay on passengers is usually articulated in terms of demand and revenue response – mindful perhaps of the financial implications of revenue compensation through the performance regimes of track access contracts. However, some previous studies have suggested that changes in performance may not directly lead to changes in demand, due to the lack of viable travel alternatives, particularly in the short-run (e.g. Batley et al., 2011). That said, whilst the demand/revenue impacts of performance may be limited by these situational constraints, this does not obviate the possibility that late running is detrimental to passenger satisfaction.

Passenger satisfaction data has been widely used in transport as there is an abundance of literature looking at how it is impacted by different journey aspects (for reviews see De Vos et al., 2013; De Oña and De Oña, 2015; Gao et al., 2018; Rong et al., 2022). As suggested by Brons and Rietveld (2009), analysis of passenger satisfaction levels and the relative importance of different aspects of the journey can usefully inform priorities for journey quality improvements. Significant heterogeneities in satisfaction levels have been reported between different types of travellers and transport mode users (e.g. Brons and Rietveld, 2009; St-Louis et al., 2014; Susilo and Cats, 2014; Transport Focus, 2015; Lunke, 2020). However, in most of the studies, travel time, value for money, performance and journey comfort were found to be the strongest determinants of passenger satisfaction.

Chapter 4 discussed the design of the currently operating compensation scheme as well as the determinants of passenger engagement with the scheme and its impacts on operators' revenues. It was highlighted that the currently used delay thresholds where passengers become eligible to claim compensation were set arbitrarily and to better understand what lengths of delays are of detrimental impact to passenger satisfaction, more research is needed - especially looking at the potential non-linearities in delay impacts. Chapter 6 noted that there is a very limited number of studies investigating the relationship between the perception of unreliability and satisfaction. Moreover, even fewer studies are related to public transport users (e.g. Transport Focus, 2015; Carrel et al., 2016; Gao et al., 2018; Monsuur et al., 2021). In all these cases, as expected, the lower performance of public transport was, however, suggested to negatively impact upon travellers' satisfaction. One of the conclusions from the analysis of perception, however, was that some shorter delays have a very low probability of being perceived, what might indicate on the possible limited consequential impacts on passenger satisfaction too. It was implied that for a delay to have an impact on satisfaction, it would typically need to be perceived first.

On the other hand, it was noted that there generally is an abundance of studies investigating the determinants of traveller satisfaction. The modelling approaches utilised by these studies are often very different as these depend on the nature of the dependent (satisfaction) variable and the availability and nature of explanatory variables. For example, the satisfaction variable may be related to a specific journey experience (e.g. Gao et al., 2018; Soza-Parra et al., 2019; Monsuur et al., 2021) or general satisfaction with transport (e.g. Cats et al., 2015; Efthymiou et al., 2019). Moreover, as suggested by Gao et al. (2018), most of the research on journey satisfaction is empirically-driven where the choice of the functional relationship is made at the discretion of researchers.

It was further noted that there is substantial precedent for exploiting satisfaction data in public policy research. In the context of the Dutch railways, Brons and Rietveld (2009) analysed passenger satisfaction with different journey aspects and their relative importance. Travel time reliability was found to be the second-worst scored journey aspect. Matching NRPS satisfaction data to operational data, Monsuur et al. (2021) found that the probability of being satisfied with a train journey decreases sharply after 30 minutes of delay (or 10 to 20 minutes if a passenger is standing in a crowded train), highlighting the importance of both travel time and comfort for passenger satisfaction. Transport Focus' (2015) analysis of the NRPS survey revealed that commuters are least satisfied with their journeys and most sensitive to delays. According to Transport Focus (2015), passenger satisfaction levels tend to start declining from the very first minute of lateness but decline less rapidly for business and leisure than for commute, until a tipping point (respectively 5 and 8 minutes of lateness) is reached, suggesting that smaller delays may have a smaller marginal impact on passenger satisfaction. Hence, there is some research discussing these topics, however, this chapter proposes some alternative approaches. One of the issues often faced when modelling choice of satisfaction categories may be found in their non-quantitative nature, complicating the interpretation of the results. For example, it is not immediately clear what a difference between '*very satisfied'* and *'fairly satisfied'* means for passengers' well-being and, hence, the implications of such change for policymakers, regulators or operators. In principle, the analysis forming part of this chapter aims to build on the work conducted by Monsuur et al. (2021) to increase understanding of the impact that delays have on passenger satisfaction. This will be explored by focusing on reported satisfaction with punctuality and analysing:

1) how the probability of being satisfied with punctuality changes with increasing levels of recorded delays,

2) the delay lengths that are detrimental to passenger satisfaction and

3) the potential indirect impact of other journey aspects, i.e. journey length and comfort on how delays affect passenger satisfaction.

In doing so, the NRPS data described in Chapter 5 is employed to conduct analysis of the relationship between recorded lengths of delays and reported satisfaction. The major contribution of the work conducted as part of this chapter compared to the numerous studies looking at passenger satisfaction lies in studying the relationship between recorded delay lengths for a given journey experience and reported satisfaction (i.e. similar to Monsuur et al., 2021). The major differences between this study and Monsuur et al. (2021) are related to:

1) The choice of the dependent satisfaction variable relating directly to satisfaction with punctuality as a specific journey aspect rather than focusing on overall journey satisfaction where performance is one of the many components affecting passenger satisfaction.

2) Apart from modelling passenger satisfaction using ordered logit model, some modifications are made to the NRPS dataset to study the relationship between performance and satisfaction using the binary choice framework (i.e. satisfaction versus dissatisfaction).

3) Aggregating the satisfaction data at the origin-destination pair level and estimating models of passenger satisfaction at an OD pair level to facilitate the application of the estimated results in policymaking (e.g. setting performance targets).

Having considered the different possible representations of the satisfaction variable, this chapter will also address two additional lines of analysis:

1) comparison of the relationship between delay length and probability of perceiving a delay versus being dissatisfied with it and
2) investigation of the marginal (dis)utility of lateness, looking at how the additional impacts of delays on satisfaction (utility) change with increasing delays.

The remainder of this chapter is structured as follows:

- Section 7.2 describes the methodology used in this chapter and the main differences between this work and the study by Monsuur et al. (2021).

- Section 7.3 then reports the results of the modelling approaches introduced in the methodology section.
- Summary of the results is presented in section 7.4.
- Section 7.5 reconciles the concepts of delay perception and satisfaction to compare the delay length thresholds where these become perceived by passengers and such that start having detrimental impacts on passenger satisfaction.
- Investigation of the potential non-linearities in delay impacts forms part of section 7.6.
- Finally, section 7.7 provides conclusions based on the work presented as part of this chapter.

7.2. Methodology

Modelling of passenger satisfaction is explored using the responses to the questionnaire from NRPS matched to operational data (as described in section 5.2) with the aim of analysing the impact of recorded performance on satisfaction reported on a 5-point Likert scale. There are two variables that describe passenger satisfaction that are particularly relevant to this work:

1) overall passenger satisfaction (Figure 27) that is assumed to be affected by delays, but also other journey characteristics (e.g. comfort) as used in Monsuur et al. (2021) and

2) satisfaction with punctuality (Figure 28) enabling studying a direct relationship between delays and satisfaction.

There are benefits and disadvantages of using each of the two variables described above. The first approach allows to study the relative importance of different journey aspects (such as the ones related to train, station, staff, ticket prices, etc.) for passenger satisfaction. This is, however, the variable that is most similar to the ones used in the previous studies of passenger satisfaction (e.g. Monsuur et al., 2021). In the case of punctuality satisfaction, there is, however, no need for controlling for these aspects. For example, it is not expected that journey quality directly affects punctuality satisfaction. Rather, it can have a complementary impact on how delays affect punctuality satisfaction – i.e. better journey quality may limit the negative impact of delays on punctuality satisfaction.

Statistical models are developed to explore the existence of a systematic and quantifiable relationship between passenger rail performance and passenger satisfaction with punctuality, whilst also controlling for the possible complementary impacts of other factors. Two alternative modelling approaches are employed:

- the ‘passenger’ model concerned with the impact of a given delay episode on a passenger’s satisfaction and
- the ‘OD’ model concerned with the impact of average performance on average levels of passenger satisfaction by origin-destination (i.e. station-to-station) flow.

Moreover, the modelling approach employs two alternative interpretations of the 5-point satisfaction scale used in the NRPS questionnaire:

- binary response where the original 5-point scale is consolidated into 2 – so as to focus on the threshold at which passenger is satisfied versus dissatisfied and
- ordered response retaining the original 5-point scale.

A priori, it is expected that delays have a negative impact on journey satisfaction. Table 17 showed the distribution of satisfaction scores for the travellers with a matched delay and how average recorded delay and overall satisfaction (NRPS question 16 in Figure 27) correspond to satisfaction with punctuality (NRPS question 9 in Figure 28). As expected, the passengers who scored their satisfaction with punctuality lower were typically subjected to lengthier delays – from around 4 minutes of average recorded delay for passengers ‘very satisfied’ with punctuality to 12 minutes for those who were ‘very dissatisfied’. Overall satisfaction levels decrease with both increasing delays and decreasing satisfaction with punctuality – from 4.6 for passengers ‘very satisfied’ with punctuality to 2.5 for passengers ‘very dissatisfied’ with punctuality as in line with the previous literature, performance is expected to be playing a key role in determining passenger satisfaction.

While the NRPS employs a 5-point satisfaction scale with the choice of satisfaction categories being:

1) very dissatisfied,
2) fairly dissatisfied,
3) neither satisfied nor dissatisfied,
4) fairly satisfied and
5) very satisfied,

the satisfaction scale gives a degree of insight into the strength of feeling, but the scale itself is not strictly cardinal. Thus, a 1-point increase on the scale (e.g. from ‘very dissatisfied’ to ‘fairly dissatisfied’) may not imply the same increase in punctuality satisfaction as an increase from ‘fairly satisfied’ to ‘very satisfied’. In this case, logistic regression would, in principle, be a more appropriate functional form than linear regression, since it takes account of such non-linearity. However, it has also been suggested that results from linear and logistic regression models applied to satisfaction data are often similar (i.e. Ferrer-i-Carbonell and van Praag, 2002; Stutzer and Frey, 2008). The main disadvantage of using logistic regression in satisfaction modelling is its complexity, as linear forms are easier to implement and interpret.

Now, consider the two alternative interpretations of the satisfaction scale – firstly the raw ordinal scale, and secondly translation of the ordinal scale into a dichotomous ‘satisfied’ or ‘not satisfied’ variable. It is noted that the binary representation of the satisfaction variable allows easier representation of the delay impacts as there is a clear threshold of satisfaction and dissatisfaction, which lends itself to ready interpretation. However, retaining the original scale allows full investigation into the dynamics of the studied relationship and can provide additional insights into the strength of satisfaction or dissatisfaction. Hence, both interpretations are utilised as explained further below.

Interpretation a): the ordinal data is converted to binary

The proposition behind this interpretation is that the 5-point ordinal scale is not a continuous scale and that at some point on the scale, there is a threshold of satisfaction versus dissatisfaction. Dichotomisation of ordinal data is a common approach in the literature, for example in medical research (Capuano et al., 2007; Sankey et al., 1998). While it is noted that dichotomisation of responses leads to a loss of information as ordinal or continuous variables provide more accurate insight into the strength of response, binary representations may help provide more direct conclusions that are easier to understand for the wider audience (DeCoster et al., 2009; Farrington & Loeber, 2000). The loss of power may be a significant concern, especially when sample sizes are small or events rare, which may often be the case with medical research. For example, in clinical trials, the distribution of outcomes may be skewed and the strength of response may also be of importance (Ceyisakar et al., 2021; Manor et al., 2000). Nonetheless, any dichotomisation needs to have a theoretical justification – for example, dichotomisation of variables simply based on the median split is considered inappropriate (Fitzsimons, 2008). As suggested, the loss of power resulting from the dichotomisation of ordinal data may be justified in the cases with specific cut-off points, such as vaccine efficacy against infection (Capuano et al., 2007). It is argued that in the case of passenger satisfaction, such a cut-off point is the switch point between satisfaction and dissatisfaction. After discussions with the railway industry, it has become clear that a question often asked by stakeholder relates to the threshold values of delays, i.e. to determine levels of delays that are acceptable for passengers. Given the ability to define the clear cut-off points and a relatively large sample size, the ordinal scale has been converted into a binary outcome, representing satisfaction versus dissatisfaction and analysed in addition to the original ordinal representation.

That said, there is more than one way to consolidate a 5-point scale into a binary variable as the satisfaction/dissatisfaction cut-off point may be interpreted in more than one way, and the following three versions are therefore tested:

- V1: 1 if passenger was 'very satisfied' with punctuality (46.3%) and 0 if otherwise (53.7%),
- V2: 1 if passenger was 'very' or 'fairly satisfied' with punctuality (79.0%) and 0 if otherwise (21.0%) and
- V3: 1 if passenger was not 'fairly' or 'very dissatisfied' with punctuality (86.3%) and 0 if otherwise (13.7%).

In what follows, these alternative versions of the binary variable are used to explore the relationship between the length of delay and the probability of being satisfied. V2 is conceptually the closest to representing satisfaction versus dissatisfaction and it also follows the usual convention adopted by Transport Focus when reporting passenger satisfaction, where the top two categories (i.e. 'very' and 'fairly satisfied') are merged and compared to the other categories (e.g. Transport Focus, 2020a).

Interpretation b): the ordinal data is retained

The proposition behind this interpretation is that the 5-point scale is a continuous ordinal scale.

In what follows, the ordinal data is modelled to discern patterns of switching between different satisfaction/dissatisfaction categories by the length of delay. Whilst the 5-point scale is rather less amenable to policy work, this is the natural form of the data and reveals additional insights, which cannot be discerned from the binary formulation.

On the basis of restricting the dataset as discussed in Chapter 5 (see Table 21 for descriptive statistics), the sub-sample taken forward comprised 72,363 'on-time' and 74,651 'delayed' responses – although for some of the models, the dataset was subject to further (albeit modest in most cases) attrition due to missing data in respect of covariates.

Having previously noted that the NRPS data has been used to study the impact of delays on satisfaction (e.g. Monsuur et al., 2021), the following section aims to discuss the similarities and differences between the two bodies of work.

7.2.1. Differences in the approach to modelling satisfaction used by Monsuur et al. (2021)

During the course of this PhD, a study was published by Monsuur et al. (2021), attempting to model passenger satisfaction and investigating how it changes with increasing delays. This study has been referred to in multiple places in the thesis as the methodology and the findings are relevant and show many similarities. The purpose of this section is to describe the main differences between the two studies as summarised in Table 32.

Table 32. Main differences between the satisfaction modelling employed in this chapter and Monsuur et al., 2021

Difference	Expected impact
Data from 10 waves of NRPS survey used as compared to 2 in Monsuur et al. (2021).	Since the number of observations is larger, this is likely to increase the accuracy and statistical significance of the estimated results.
Monsuur et al. (2021) restricts the dataset to responses where delay was both recorded and reported. Such restriction is not imposed in this chapter.	Retaining responses where passenger did not perceive a delay or was not delayed aims to allow for studying the impact of both perceived and unperceived delays on passenger satisfaction.
The core analysis in this chapter uses data on satisfaction with punctuality rather than overall satisfaction used in Monsuur et al. (2021)**.**	Satisfaction with punctuality is expected to be more directly linked to a delay experience. Therefore, it allows understanding the impacts of delays on satisfaction and how these are affected by journey quality and length. This is somewhat different to the impact that journey quality has on overall satisfaction.
The recorded delays were retrieved using TRUST database in Monsuur et al. (2021) instead of HSP database (used for this work).	As noted by Monsuur et al. (2021), TRUST [Trains Running Under System TOPS (Total Operation Processing System)] records real-time train performance by comparing recorded train timings at designated timing points (e.g. stations, junctions). HSP only provides the timetabled and recorded timings at timetabled stopping points, what is sufficient for the purposes of this study.
Apart from the ordered logit also used by Monsuur et al. (2021), alternative approaches are also introduced in this chapter.	The different modelling approaches used in this chapter aim to decrease the complexity of the estimated models to allow easier interpretation of the results, increase the usefulness of this work for policymakers as well as serve as sensitivity tests.

As noted above, despite the similarities in the research questions investigated, there are some important differences that make the work conducted as part of this chapter unique with respect to the variables and methodological approaches used. This aims to provide new insights as well as facilitate the application of the results to policymaking/regulatory contexts, such as setting performance targets or designing compensation schemes.

Having discussed the differences in the analysis of passenger satisfaction conducted as part of this chapter and the work done by Monsuur et al. (2021), the following section aims to provide a more detailed description of the modelling approaches used as part of this chapter.

7.2.2. Modelling approach

As aforementioned, there are three dimensions of the satisfaction modelling undertaken as part of this chapter:

1) satisfaction (dependent) variable relating to overall or punctuality satisfaction,
2) binary or ordinal representation of the satisfaction variable and
3) individual passenger or aggregated OD-level modelling approach.

Further details are now given on these approaches.

7.2.2.1. The 'passenger' model

Passenger model is concerned with modelling the satisfaction reported by individual passengers following the experience of a delay event of a specified duration. This approach offers particular insight into passenger satisfaction across the distribution of delays. Within the passenger model, satisfaction with punctuality is modelled using both interpretations of the satisfaction variable (i.e. binary and ordinal), and the binary interpretation is also modelled using all three versions of the satisfaction threshold (i.e. V1-V3).

In more formal terms, the binary response variable takes the form:

$$Y = \begin{cases} 1 \; if \; passenger \; is \; satisfied \; with \; punctuality \\ 0 \; if \; passenger \; is \; not \; satisfied \; with \; punctuality \end{cases} \tag{18}$$

The binary representation is modelled in a way that is methodologically very similar to modelling delay perception in Chapter 6 with the initial model conforming to the specification of model 1 in 6.4.2 and the extended version conforming to model 3. The only difference is the change of the outcome variable.

The binary response models (i.e. pertaining to the binary formulation of the data), as well as the ordered response models (i.e. pertaining to ordinal data) to follow, were estimated using Stata 17 (StataCorp, 2021). The estimates are indicative of the direction and strength of relationships between satisfaction with punctuality and the explanatory variables but may be challenging to interpret directly. Similarly as in the case of delay perception, in addition to the estimated coefficients, predicted probabilities are shown graphically for a range of arrival delay lengths from 0 to 30 minutes (for binary models) with comparisons made between the predicted delay dissatisfaction threshold (at p=0.5 of satisfaction probability), i.e.

$$0.5 = \frac{1}{(1 + e^{-(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)})}$$

(19)

Following the analysis outlined above, the binary response model was extended to estimate passengers' probability of choosing each of the five satisfaction levels. As the dependent variable can now take one of the five categories (more similar to the methodology used by Monsuur et al., 2021), which are in sequential order, an ordered logit model (McCullagh, 1980) was employed to estimate the latent continuous variable y^*. In this case, the probability of choosing a satisfaction category *i* is estimated for *k* thresholds, thus:

$$P(Y = i) = P(k_{i-1} < y^* \leq k_i)$$

(20)

In this case, the probability of $Y = i$ is:

$$P(Y = i) = F(k_i - \eta) - F(k_{i-1} - \eta)$$

(21)

where:

F is the cumulative distribution function of the logistic distribution and $y^* = \eta + \epsilon$.

Similarly, to the binary outcome models introduced above, the choice of explanatory variables conforms to the models of delay perception, i.e. model 1 for the initial model:

$$\eta = \sum_{i=1}^{i=6} (JP_i \times \beta_{1,i} + JP_i \times \beta_{2,i} \times L_{A,i}^{+})$$

(22)

and model 3 for the extended model:

$$\eta = \sum_{i=1}^{i=6} (JP_i \times \beta_{1,i} + Seat_i \times JP_i \times \beta_{2,i} \times L^{+}_{A,i} + Seat_i \times JP_i \times \beta_{3,i} \times L^{+}_{A,i} \times SJT_i + JP_i \times \beta_{4,i} \times L^{+}_{A,i} \times L^{+}_{D,i}) \tag{23}$$

where:

JP_i is a journey purpose dummy variable for each of the 6 journey purposes that takes the value of 1 when it matches the respondents' journey purpose or 0 otherwise

$L^{+}_{A,i}$ is the delay length at arrival which is defined as the difference between the actual and scheduled arrival for all cases where the difference is positive; when the difference is negative, such responses are treated as on-time arrival

SJT_i is the scheduled journey time

$Seat_i$ is a dummy variable that takes the value of 1 if passenger reported having a seat or 0 otherwise

$L^{+}_{D,i}$ is the delay length at departure which is defined as the difference between the actual and scheduled departure. If a train departed before its scheduled departure time, this is counted as on time departure.

In logistic regression, the coefficients represent the change in log-odds of the outcome (i.e. being satisfied with punctuality) for a given change in explanatory variables, indicating the strength and direction of the relationship. This highlights one of the limitations of logistic regression relative to linear regression, in that the direct interpretation of the coefficients is less straightforward. However, the estimated coefficients can be converted into margins, representing the predicted outcome for a given change in an explanatory variable. Nevertheless, in the case of multiple explanatory variables, and especially if these are continuous, such interpretation becomes difficult because of the multi-dimensionality of the problem. This is further complicated in the ordered logit model where there are always five outcome categories leading to a set of five estimated probabilities for each set of values of explanatory variables. Hence, in the case of the ordered logit model in its extended form, this limits the ability to present the results of the modelling graphically.

7.2.2.2. The 'OD' model

The main difference between passenger and OD models is that, whilst the former consider the impacts of a delay event of a given length on journey satisfaction at the individual passenger level, the latter consider average performance across the responses gathered on a given OD pair and its impacts on the proportion of satisfied passengers.

This reformulation of the problem necessitates averaging of the raw passenger-level data for each OD pair with the dataset restricted to the OD pairs with at least 25 responses (arbitrary threshold) – which in the present work markedly reduced the scale of the dataset to 676 observations (OD pairs). It follows that the proportion of satisfied passengers (again using the three alternative versions of delay satisfaction V1-V3) is now expressed in relation to Average Passenger Lateness (APL). Similarly, the other variables represent means or proportions – for example, seat availability is presented as the proportion of passengers who were able to find a seat for a given OD. The averages are not weighted for flow sizes as the distribution of delays in NRPS is not necessarily representative of that in the network.

The model is formulated in terms of the proportion of passengers satisfied with punctuality, which lends itself to a fractional outcome logit regression (Papke & Wooldridge, 2008) – as stated below. The model is conceptually similar to the binary logit model, but rather than considering the probability of being satisfied, here we consider the proportion of satisfied passengers. In more formal terms:

$$E(Y|X) = F(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)$$

(24)

where:

$$F = \frac{1}{(1 + e^{-(\beta_o + \beta_1 X_1 + \cdots + \beta_i X_i)})}$$

(25)

To simplify the studied relationship, the interaction between delay length at arrival and departure was excluded and journey purpose subcategories were grouped together, hence:

$$E(Y|X) = F(\sum_{i=1}^{i=3} (JP_i \times \beta_{1,i} + Seat_i \times JP_i \times \beta_{2,i} \times L^{+}_{A,i} + Seat_i \times JP_i \times \beta_{3,i} \times L^{+}_{A,i} \times SJT_i)$$

Unlike the passenger model, the OD model is not amenable to ordinal data, since such data does not readily lend itself to averaging.

7.3. Modelling results

This section reports the results for the modelling approaches described in the previous section, starting with the so-called passenger model (and its binary and ordinal representations) and subsequently, the OD model using only the binary representation of the response variable).

The following models are estimated, as summarised in Table 33:

1) **Passenger model**

   a) **Binary**

      i. Initial model for three versions of punctuality satisfaction (Table 34)

      ii. Extended model for three versions of punctuality satisfaction with additional control variables (Table 36)

   b) **Ordinal**

      iii. Initial model for punctuality satisfaction (Table 39)

      iv. Initial model for overall satisfaction (Table 39)

      v. Extended model for punctuality satisfaction with additional control variables (Table 41)

2) **The OD model**

   a) **Binary only**

      i. Extended model with control variables (Table 45)

Table 33 Guide to different model permutations estimated (OS refers to overall satisfaction while PS refers to satisfaction with punctuality)

	Passenger model		**OD model**
	Binary	**Ordered**	**Binary**
Initial	Table 34 (PS)	Table 39 (PS, OS)	-
Extended	Table 36 (PS)	Table 41 (PS)	Table 45 (PS)

7.3.1. The 'passenger' model

Binary logit

The binary logit passenger model estimates the probability of a passenger being satisfied with punctuality, having experienced a journey, which may have been on-time or late –

with satisfaction being treated as a binary rather than ordinal outcome. The probability of being satisfied was modelled using each of the three versions of satisfaction (V1-3), outlined in section 7.2, for different journey purposes using the initial and then extended model specification.

In the first model (reported in Table 34), the binary outcome for the three versions of satisfaction (V1-3) is explained by recorded delay at arrival, whilst allowing varying impacts by journey purpose. All the coefficients are statistically significant and, as expected, arrival delay has a negative impact on satisfaction with punctuality, meaning that the probability of being satisfied decreases with increasing delay length. McFadden's R^2 values of around 0.1-0.2 are reported, which is of a similar magnitude to the models of delay perception. This suggests that while a general relationship is well-described by the estimated models, the models are unable to predict the choice of some of the respondents, e.g. where satisfaction was reported despite a long delay or dissatisfaction was reported when a relatively small delay was recorded.

The purpose of the constant is to capture the probability of being satisfied with punctuality under 'no delay'. Intuitively, one might expect passengers to report complete satisfaction with punctuality if they are not subjected to late running. However, significant differences in the constant were found between different types of passengers, in particular suggesting that commuters' probability of being satisfied is lower in the 'no delay' case than that for leisure and business travellers. When passengers reported arriving on time, commuters' average satisfaction with punctuality was found to be 3.66, as compared to 4.38 for the whole dataset. By contrast, when passengers reported arriving late, the respective average satisfaction scores were 2.93 and 3.08. This suggests that commuters are unlikely to be 'very satisfied' with punctuality regardless of the level of performance. The possible reasons for this may include strategic bias where passengers score satisfaction lower to influence decision-making, or passengers reflecting more generally on performance rather than describing their satisfaction with the specific journey leg (as asked in the survey).

Table 34 Estimated binary logistic regression coefficients for the three versions of the binary representation of the punctuality satisfaction variable

	Sat_V1	Sat_V2	Sat_V3
Constant	0.468***	2.554***	3.374***
	(24.99)	(79.56)	(76.10)
BnL	-0.154***	-0.304***	-0.387***
	(-4.97)	(-6.21)	(-5.92)
CL	-1.294***	-1.505***	-1.690***
	(-49.39)	(-41.04)	(-34.58)
CnL	-1.312***	-1.610***	-1.847***
	(-41.99)	(-40.71)	(-35.95)
LF	0.194***	-0.0271	-0.0388
	(7.82)	(-0.65)	(-0.67)
LR	0.292***	0.188***	0.135**
	(13.29)	(4.95)	(2.58)
Arrival delay			
BL	-0.0954***	-0.133***	-0.154***
	(-25.78)	(-34.55)	(-35.66)
BnL	-0.138***	-0.175***	-0.183***
	(-23.64)	(-32.78)	(-33.22)
CL	-0.184***	-0.171***	-0.178***
	(-25.86)	(-39.87)	(-42.89)
CnL	-0.211***	-0.180***	-0.181***
	(-21.42)	(-32.02)	(-34.11)
LF	-0.135***	-0.163***	-0.178***
	(-31.52)	(-38.64)	(-38.77)
LR	-0.109***	-0.152***	-0.163***
	(-48.15)	(-63.14)	(-61.51)
N	133478	133478	133478
Log-likelihood	-81605.2	-53501.1	-39138.9
Pseudo R^2	0.117	0.170	0.200

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced

Figure 44 shows how the predicted probabilities of delay satisfaction change with increasing delays. For better comparison between the different versions of delay satisfaction (i.e. V1-3), as well as across different journey purposes, Table 35 shows the predicted delay lengths where the probability of accepting a delay is equal to 0.5 (i.e. the threshold at which a passenger becomes more likely to be dissatisfied than satisfied). As expected, the probability of being 'very satisfied' with punctuality (i.e. Version 1) is always smaller than the probability of being satisfied under the more relaxed definitions of delay satisfaction (i.e. Versions 2 and 3), and this applies to all journey purposes. The large difference between the probabilities of accepting a delay under V1 vs. V2-3 of the response variable suggests that the definition of satisfaction versus dissatisfaction has a large impact on the results.

Commuters' probability of being satisfied with punctuality never reaches 0.5 under V1 of delay satisfaction. However, for business travellers, the estimated probability of being satisfied with punctuality is larger than 0.5 for delays smaller than 5 minutes (London travellers) and 2.3 minutes (non-London travellers). In the case of leisure travellers, the threshold is 4.9 minutes for travellers on Full fares and 7.0 minutes on Reduced fares.

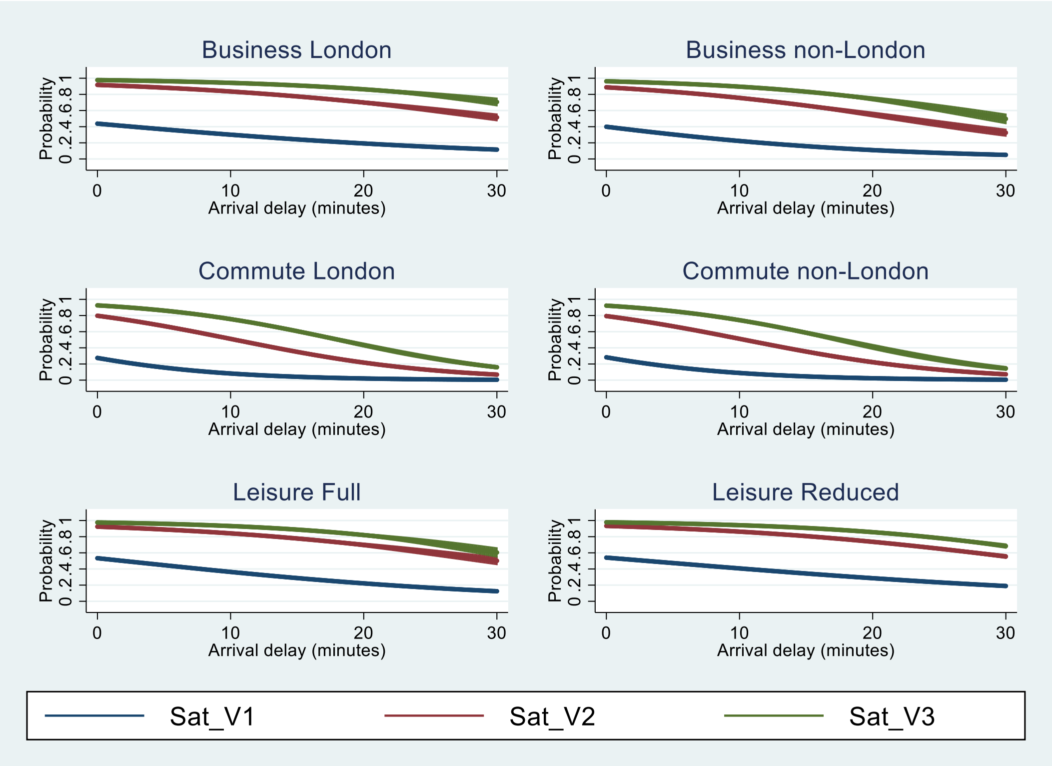


Figure 44 Probability of 'delay satisfaction' for increasing delay lengths and different journey purposes based on the three definitions of delay satisfaction V1: (5) vs (1-4) ; V2: (4-5) vs (1-3); V3: (3-5) vs (1-2) based on the model from Table 34.

Under V2 with a more relaxed definition of delay satisfaction (i.e. ‘very satisfied’ and ‘fairly satisfied’ categories), the predicted delay lengths where passengers are more likely to be dissatisfied with punctuality (i.e. p=0.5) increase. For business travellers, the threshold increases to 19.3 minutes (London travellers) and 12.9 minutes (non-London travellers). The respective thresholds for commuters are much lower – 6.2 minutes for London and 5.3 minutes for non-London travellers. The thresholds for leisure travellers are more similar to those of business travellers – 15.5 minutes for Full fare and 18.1 for Reduced. The relatively large difference between London and non-London business travellers may be a result of differences in journey quality, slightly longer journeys (though it is noted that the average journey length is only slightly longer for London business travellers) or inclusion of larger buffers which make these travellers less sensitive to any potential delays.

The estimated thresholds suggest that, as expected, commuters are the most sensitive group of travellers with respect to delays. Relaxing the definition of delay satisfaction further (i.e. inclusion of the ‘neither satisfied nor dissatisfied’ category) increases the predicted delay length satisfaction thresholds by around 3 minutes for all types of passengers.

Table 35 Delay length thresholds at the estimated probability p=0.5

	Delay at p=0.5 V1	**Delay at p=0.5 V2**	**Delay at p=0.5 V3**
Business London	5.0	19.3	21.9
Business non-London	2.3	12.9	16.4
Commute London	-	6.2	9.5
Commute non-London	-	5.3	8.5
Leisure Full	4.9	15.5	18.8
Leisure Reduced	7.0	18.1	21.6

The comparison presented above asks for a commentary regarding how the estimated satisfaction thresholds compare with the perception thresholds estimated in Chapter 6 what will be discussed in section 7.5.

The V2 of the model is now re-estimated with the addition of control variables (as in the case of the perception models) with the results presented in Table 36. The estimated model now has a larger number of dimensions as delay satisfaction is explained by multiple continuous variables and interactions.

Table 36 Estimated coefficients for the binary logit model with controls (aligning with the model 3 of delay perception estimated in Chapter 6)

	(V2)	(V2a)
Constant	2.662***	1.771***
	(62.67)	(6.33)
BnL	-0.397***	-0.239
	(-6.18)	(-0.65)
CL	-1.516***	-0.701*
	(-31.36)	(-2.38)
CnL	-1.672***	-1.121***
	(-32.02)	(-3.67)
LF	-0.0902	0.247
	(-1.62)	(0.77)
LR	0.0573	0.166
	(1.16)	(0.53)
Seat=1		
BL		1.042***
		(3.60)
BnL		1.026***
		(4.14)
CL		0.482***
		(4.61)
CnL		0.686***
		(5.06)
LF		0.792***
		(4.74)
LR		0.899***
		(6.07)
Arrival delay (Seat=0)		
BL	-0.311***	-0.114*
	(-5.75)	(-2.21)
BnL	-0.303***	-0.104**
	(-8.26)	(-2.96)
CL	-0.279***	-0.138***
	(-11.53)	(-5.79)
CnL	-0.351***	-0.111***
	(-12.97)	(-3.76)
LF	-0.332***	-0.136***
	(-12.31)	(-5.18)
LR	-0.275***	-0.100***
	(-13.77)	(-5.32)
Arrival delay (Seat=1)		
BL	-0.171***	-0.160***
	(-14.65)	(-13.20)
BnL	-0.193***	-0.148***
	(-14.16)	(-10.70)
CL	-0.185***	-0.167***
	(-15.09)	(-12.94)
CnL	-0.228***	-0.126***
	(-15.74)	(-7.95)
LF	-0.203***	-0.146***
	(-18.39)	(-12.75)
LR	-0.161***	-0.125***
	(-24.50)	(-18.99(
Arrival delay x SJT (Seat=0)		
BL	0.00111*	0.0000748
	(2.02)	(0.12)
BnL	0.000796*	-0.000587
	(2.51)	(-1.43)

CL	-0.000453	-0.000324
	(-0.62)	(-0.45)
CnL	0.00178*	0.000588
	(1.99)	(0.58)
LF	0.00137***	0.000315
	(4.06)	(0.94)
LR	0.000648***	-0.0000399
	(3.54)	(-0.22)
Arrival delay x SJT (Seat=1)		
BL	0.000549***	0.000222*
	(6.70)	(2.31)
BnL	0.000620***	0.0000332
	(6.92)	(0.30)
CL	0.00112***	0.000755**
	(5.35)	(3.12)
CnL	0.00134***	0.000198
	(4.84)	(0.61)
LF	0.000814***	0.000154
	(8.58)	(1.38)
LR	0.000472***	-0.0000457
	(11.31)	(-0.96)
SJT (Seat=0)		
BL		0.00122
		(0.29)
BnL		0.00899*
		(2.34)
CL		-0.0114***
		(-3.77)
CnL		-0.00877
		(-1.75)
LF		0.00180
		(0.60)
LR		0.00228
		(1.20)
SJT (Seat=1)		
BL		0.00297***
		(3.38)
BnL		0.00427***
		(4.20)
CL		-0.00242
		(-1.81)
CnL		-0.000694
		(-0.35)
LF		0.00402***
		(3.98)
LR		0.00480***
		(9.58)
Departure delay		
BL		-0.307***
		(-19.14)
BnL		-0.339***
		(-20.49)
CL		-0.271***
		(-21.09)
CnL		-0.269***
		(-18.08)
LF		-0.305***
		(-23.30)
LR		-0.304***
		(-36.11)

Departure delay x arrival delay		
BL	-0.00919***	0.0103***
	(-10.37)	(10.31)
BnL	-0.00498***	0.0121***
	(-5.58)	(13.81)
CL	-0.00200*	0.0121***
	(-2.43)	(15.89)
CnL	0.00161*	0.0110***
	(2.01)	(14.83)
LF	-0.00242***	0.0117***
	(-3.65)	(16.33)
LR	-0.00621***	0.00966***
	(-15.75)	(20.89)
N	87882	87882
LL	-34114.7	-32109.5
Pseudo R^2	0.196	0.243

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced; SJT: scheduled journey time, Seat=0 represents a standing passenger

Under model V2 in its extended form, delays at both departure and arrival are suggested to have a negative impact on delay satisfaction. This highlights the potential impact that departure delay may have on how passengers perceive late running, therefore, increasing the likelihood of being dissatisfied with punctuality. Alternatively, it may be related to the additional inconvenience and uncertainty related to waiting for a delayed train on platform. Having a seat has a mitigating impact on delay satisfaction, as seated passengers may be able to use their travel time more productively. All else being equal, journey time typically has a small positive impact on delay satisfaction, especially for seated passengers. In the case of standing London commuters, scheduled journey length is suggested to have a negative impact on delay satisfaction (though not significant). To some extent, this may be explained by the suggestion by Cats et al. (2015) that long commute is generally associated with lower satisfaction with public transport.

Model V2 was also re-estimated with the inclusion of all levels of the interacted variables (model V2a). Similarly to model 3 of delay perception (from the previous chapter), in the case of the fully specified model, the model is affected by multicollinearity and the probability of being satisfied is suggested to start increasing with the length of delay at arrival for longer delays at departure, as depicted in Annex II (section B). This confirms that the fully specified model is unable to correctly capture the modelled relationship, hence, the originally proposed model (V2) is retained as the preferred one.

The predicted perception thresholds in Chapter 6 were investigated in great detail (since the amount of research concerning delay perception is much more limited). For brevity, the probabilities of being satisfied with punctuality were estimated based on the extended binary logit model (reported in Table 36) for each of the 6 journey purposes and only for the 4 scenarios represented in Table 37 (as shown in Figure 45).

Table 37 Scenarios used in reporting probabilities of being satisfied with punctuality

Scenario	SJT	Departure delay	Seat	Arrival delay
1	Mean value	0	No	
2	for each of the	0	Yes	0-30 minutes
3	journey	15	No	
4	purposes	15	Yes	

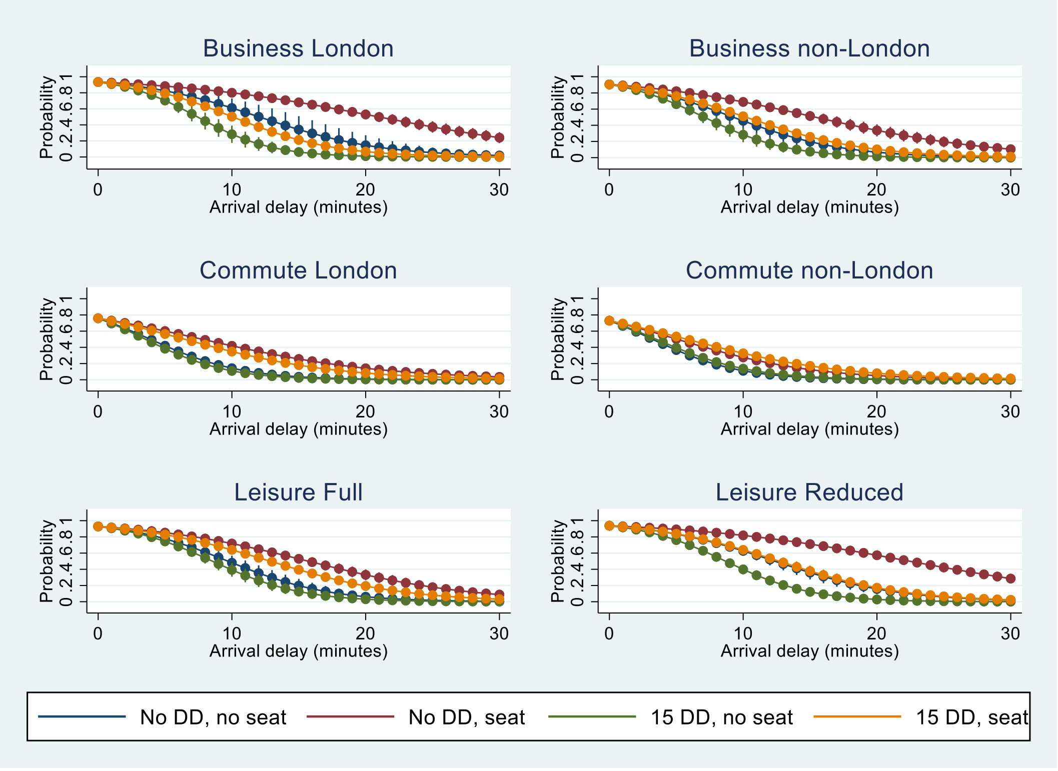


Figure 45 Probability of 'delay satisfaction' for increasing delay lengths and scenarios shown in Table 37

The delay length dissatisfaction thresholds (defined as the arrival delay length with equal predicted probabilities of satisfaction and dissatisfaction) were estimated (Figure 46), similarly to the delay perception thresholds (from Figure 39). The differences between the predicted dissatisfaction thresholds are mostly similar for leisure and business travellers, with much lower thresholds typically predicted for commuters. Table 38 provides a summary of the estimated values.

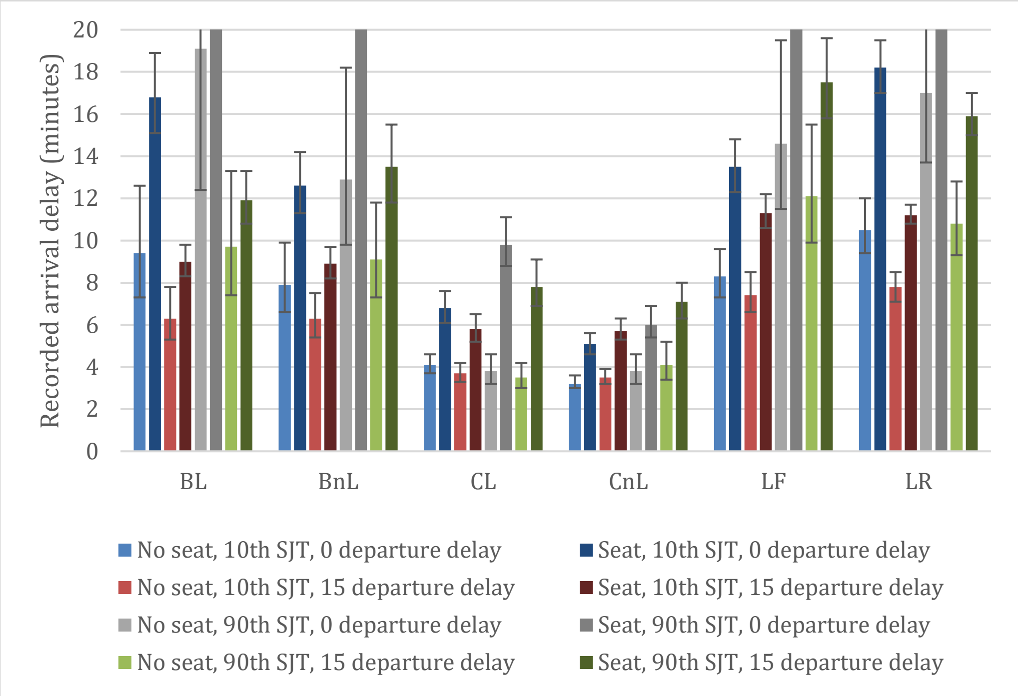


Figure 46 Estimated thresholds of delay dissatisfaction

Table 38 Summary of the impacts of having a seat, journey lengths and departure delay on arrival delay dissatisfaction thresholds

Journey purpose	Minimum threshold (i.e. standing, long departure delay, short journey)	Increase in the threshold due to		
		Having a seat	Longer journey	No delay at departure
Business	6.3	+2 to +11	+3 to +13	+2 to +20
Commute	3.5	+2 to +6	0 to +3	-1 to +2
Leisure	7.4	+3 to +15	+3 to +12	+1 to +15

The summary presented above highlights the impact of both journey quality and length on the impact of delays on passenger satisfaction. Moreover, for non-commuters, it would seem that being delayed at departure from the origin station has a large impact on delay satisfaction, likely due to the additional stress, uncertainty and discomfort related to late departure as well as any possible impacts that delay at departure may have on the perception of final performance. Section 7.5. will compare the concepts of delay perception and satisfaction.

This section presented results from the models of passenger satisfaction that used the binary representation of the data. The subsequent section utilises the original (ordinal) nature of the satisfaction scale, estimating ordered logit models of passenger satisfaction.

Ordered logit

The ordered logit works with the natural form of the data by modelling the probability of choosing each of the five original satisfaction categories for a given delay length. This allows insight into the dynamics of the reported satisfaction, especially with respect to how the most likely satisfaction category changes with increasing delay length.

Firstly, an ordered logit model is estimated in its initial form. The choice of satisfaction categories is explained by arrival delay while allowing for differing impacts by journey purpose. Of note, the models are estimated for two dependent variables – punctuality satisfaction in the first model (as in the case of the previous models) and overall satisfaction in the second model (as in the case of Monsuur et al., 2021) with the estimated coefficients presented in Table 39. It is noted that satisfaction with a journey is likely affected by more factors, such as those related to journey quality. However, as modelling of overall journey satisfaction is not the main focus of the thesis, more sophisticated models of overall journey satisfaction are not introduced in this chapter with the overall journey satisfaction model reported for reference only.

Table 39 Estimated logistic regression coefficients for the ordered logit model of punctuality and overall satisfaction

	(1 Punc_Sat)	(2 Overall_Sat)
Journey purpose		
BnL	-0.178***	-0.187***
	(-6.19)	(-6.80)
CL	-1.360***	-0.805***
	(-59.27)	(-36.18)
CnL	-1.450***	-0.798***
	(-55.56)	(-31.21)
LF	0.157***	0.331***
	(6.64)	(14.91)
LR	0.280***	0.378***
	(13.37)	(19.21)
Arrival delay		
BL	-0.114***	-0.0648***
	(-37.84)	(-21.99)
BnL	-0.159***	-0.0911***
	(-41.65)	(-24.48)

CL	-0.163***	-0.135***
	(-51.32)	(-45.71)
CnL	-0.169***	-0.136***
	(-42.15)	(-36.77)
LF	-0.155***	-0.0782***
	(-46.71)	(-24.35)
LR	-0.134***	-0.0653***
	(-71.93)	(-35.98)
Threshold 1	-4.186***	-4.618***
	(-181.24)	(-170.56)
Threshold 2	-2.994***	-3.293***
	(-148.88)	(-160.26)
Threshold 3	-2.354***	-2.194***
	(-122.67)	(-120.02)
Threshold 4	-0.518***	0.192***
	(-29.09)	(11.38)
N	133478	137176
LL	-148630.6	-148002.3
Pseudo R^2	0.0971	0.0554

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR- Leisure Full/Reduced; Punc_Sat and Overall_Sat refer to satisfaction with punctuality and overall journey satisfaction

The dynamics of the changes in the dominant categories are different for the models with the dependent variable being satisfaction with punctuality (Figure 47) and overall journey satisfaction (Figure 48). It is worth making some references to the models estimated by Monsuur et al. (2021) where the overall satisfaction model contained a larger number of explanatory variables. Monsuur et al. (2021) estimated that the dominant category choice of overall satisfaction is 'fairly satisfied' up to around 50 minutes where it changes to 'very dissatisfied'. 30 minutes was indicated to be the delay length that is detrimental to passenger satisfaction. The analysis conducted as part of this chapter suggests a similar relationship, but a much quicker change between satisfaction and dissatisfaction, especially so for commuters. While the exact comparisons are difficult due to the differences in the modelling approaches or variables used, Table 40 summarises the dominant satisfaction category choice for increasing delays from the models reported in this section for overall and punctuality satisfaction as well as in Monsuur et al. (2021).

The key observation is that the differences between the two outcome variables may be explained by overall journey satisfaction also being affected by other aspects of the journey. This also suggests that especially in the context of this piece of work, using satisfaction with punctuality as a variable of interest is preferable. Moreover, the smaller difference in the predicted probabilities (i.e. for the model of punctuality and overall satisfaction) for commuters would suggest that these travellers' overall satisfaction is much more impacted by the delay lengths (with a much smaller impact of other journey aspects).

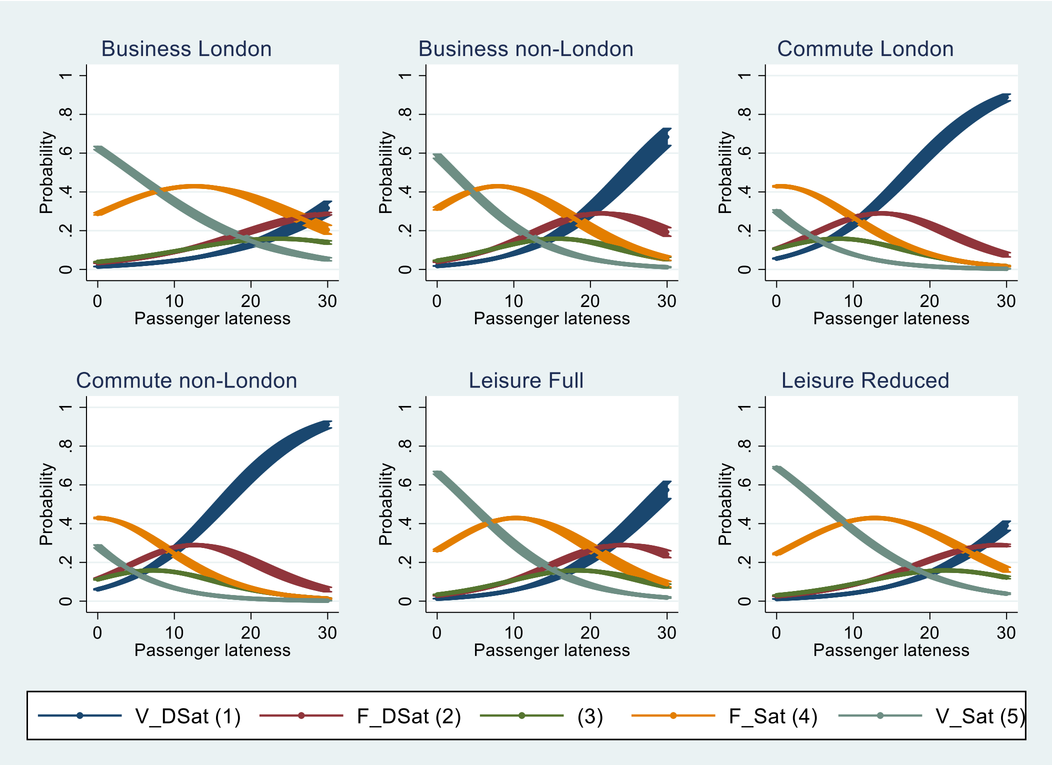


Figure 47 Probability of punctuality satisfaction for increasing delays and different journey purposes

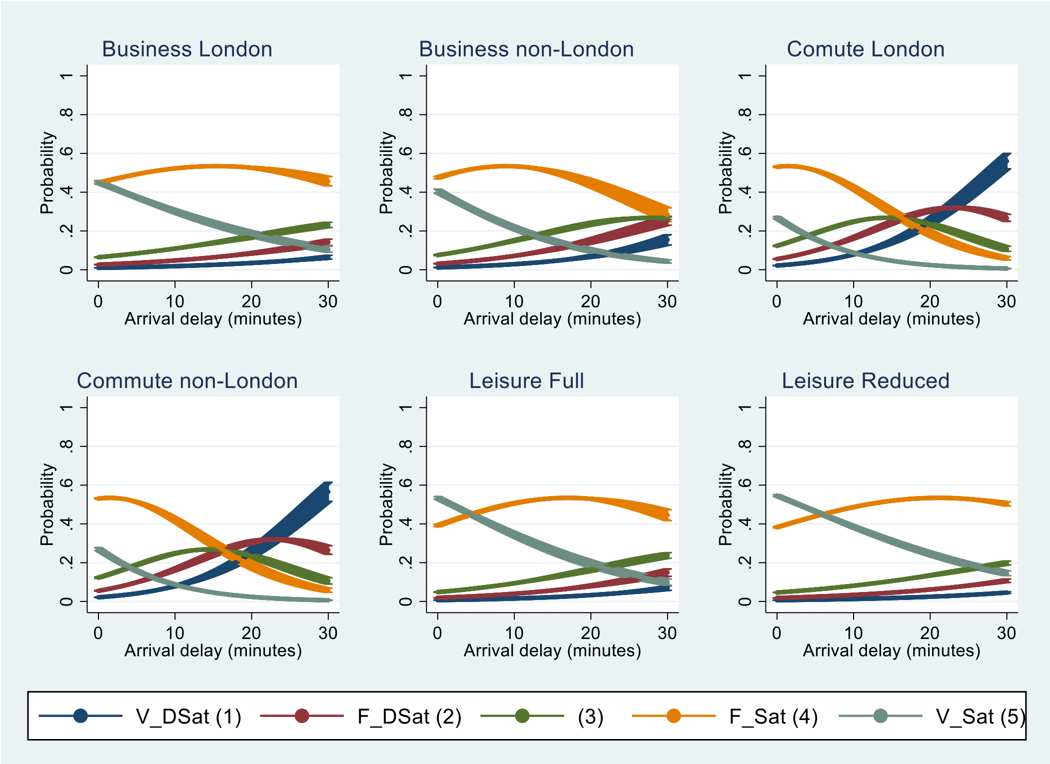


Figure 48 Probability of overall satisfaction for increasing delays and different journey purposes

Table 40 Comparison of the estimated dominant satisfaction choice based on overall or punctuality satisfaction and Monsuur et al. (2021)

Journey purpose	Punctuality Satisfaction	Overall Satisfaction	Monsuur et al. (2021)
Business	Very satisfied dominant up to 10 minutes Fairly satisfied dominant between 10 to 20-25 minutes Very dissatisfied dominant for 20-25+ minutes	Fairly satisfied dominant up to 30 minutes	Fairly satisfied dominant up to 50 minutes
Commute	Fairly satisfied dominant up to 10 minutes Very dissatisfied dominant for 10+ minutes	Fairly satisfied dominant up to 15 minutes Very dissatisfied dominant for 15+ minutes	Very dissatisfied dominant for 50+ minutes

Leisure	Very satisfied dominant up to 10 minutes Fairly satisfied dominant between 10 to 20-25 minutes Very dissatisfied dominant for 20-25+ minutes	Very satisfied dominant up to 5 minutes Fairly satisfied dominant for 5+ minutes

With reference to Table 39, the ordered logit model is subsequently extended to include the same control variables as the binary response model from Table 36 with results reported in Table 41 (model Punc_Sat). In common with the binary model, the results highlight the importance of journey length and quality in determining the impacts of delays on passenger satisfaction as the estimated coefficients are of the expected signs. The extended model of punctuality satisfaction has a slightly higher pseudo R^2 of over 0.1. While this is somewhat lower than the values reported in Monsuur et al. (2021) for the ordered models of overall journey satisfaction, this is probably due to the more complex nature of punctuality satisfaction as opposed to overall journey satisfaction. The estimated margins are not presented for the ordered logit model in its extended version due to its complexity as the choice probabilities for each of the scenarios relate to five, not two choices (as is the case with the binary outcome models). In practical terms, the same amount of information provided for the binary logit model in Figure 45 would need to be shown on four different plots for the ordered logit model, which also highlights the benefits of the binary representation of the modelling framework.

The model was re-estimated with the inclusion of all levels of interacted variables (model Punc_Sat_1). Similarly as with the binary logit models, reported in the previous sections, the fully specified model predicts that at larger values of departure delay, the probability of being very dissatisfied (response 1) starts decreasing with longer delay at arrival (Annex II, section C). Hence, the fully specified model is not able to correctly predict the studied relationship. Additionally, there is the aforementioned problem with multicollinearity, hence, the original more parsimonious model is retained as the preferred model.

Table 41 Estimated ordered logit coefficients for punctuality satisfaction with controls

	(Punc_Sat)	(Punc_Sat_1)
BnL	-0.245***	-0.446
	(-6.57)	(-1.69)
CL	-1.394***	-0.946***
	(-47.34)	(-4.45)
CnL	-1.465***	-1.289***
	(-42.91)	(-5.70)
LF	0.110***	0.0348
	(3.58)	(0.15)
LR	0.187***	0.00969
	(6.99)	(0.04)
Seat=1		
BL		0.686***
		(3.38)
BnL		0.891***
		(5.00)
CL		0.409***
		(4.80)
CnL		0.685***
		(5.89)
LF		0.792***
		(7.00)
LR		0.744***
		(7.32)
Arrival delay (Seat=0)		
BL	-0.193***	-0.0838*
	(-4.66)	(-2.24)
BnL	-0.255***	-0.108***
	(-10.19)	(-4.36)
CL	-0.257***	-0.147***
	(-16.71)	(-9.22)
CnL	-0.314***	-0.118***
	(-17.06)	(-5.40)
LF	-0.284***	-0.125***
	(-14.62)	(-6.25)
LR	-0.262***	-0.104***
	(-16.99)	(-6.87)
Arrival delay (Seat=1)		
BL	-0.143***	-0.131***
	(-16.33)	(-14.36)
BnL	-0.179***	-0.131***
	(-17.76)	(-12.54)
CL	-0.180***	-0.157***
	(-18.50)	(-15.08)
CnL	-0.233***	-0.136***
	(-20.35)	(-10.49)
LF	-0.189***	-0.133***
	(-21.97)	(-14.83)
LR	-0.140***	-0.101***
	(-29.24)	(-20.62)
Arrival delay x SJT (Seat=0)		
BL	0.000114	0.00000429
	(0.27)	(0.01)
BnL	0.000682**	-0.000286
	(2.60)	(-0.97)

CL	0.000610	0.000276
	(1.44)	(0.63)
CnL	0.000263	0.0000166
	(0.41)	(0.02)
LF	0.000892***	0.000183
	(4.07)	(0.72)
LR	0.000598***	0.0000866
	(3.88)	(0.55)
Arrival delay x SJT (Seat=1)		
BL	0.000397***	0.000233***
	(6.64)	(3.34)
BnL	0.000553***	0.000119
	(8.02)	(1.48)
CL	0.000814***	0.000470*
	(4.77)	(2.35)
CnL	0.000867***	0.0000248
	(3.87)	(0.09)
LF	0.000764***	0.000269**
	(10.59)	(3.22)
LR	0.000431***	0.0000236
	(14.29)	(0.69)
SJT (Seat=0)		
BL		-0.00313
		(-1.07)
BnL		0.00660**
		(2.66)
CL		-0.0125***
		(-4.90)
CnL		-0.0136**
		(-3.10)
LF		0.000764
		(0.37)
LR		0.0000257
		(0.02)
SJT (Seat=1)		
BL		0.000921*
		(2.11)
BnL		0.00259***
		(4.81)
CL		-0.00214*
		(-2.07)
CnL		-0.00248
		(-1.56)
LF		0.00194***
		(3.88)
LR		0.00329***
		(13.39)
Departure delay		
BL		-0.264***
		(-22.96)
BnL		-0.280***
		(-23.60)
CL		-0.256***
		(-25.19)
CnL		-0.251***
		(-21.14)
LF		-0.280***
		(-28.04)
LR		-0.267***
		(-44.35)

Departure delay x arrival delay		
BL	-0.00551***	0.00783***
	(-10.65)	(11.25)
BnL	-0.00288***	0.00891***
	(-5.68)	(13.79)
CL	-0.0000687	0.0112***
	(-0.14)	(18.50)
CnL	0.00328***	0.0107***
	(6.29)	(17.97)
LF	-0.00256***	0.00956***
	(-5.23)	(16.04)
LR	-0.00493***	0.00709***
	(-20.29)	(21.17)
cut1	-4.393***	-4.013***
	(-147.00)	(-20.08)
cut2	-3.117***	-2.702***
	(-120.47)	(-13.56)
cut3	-2.458***	-2.004***
	(-99.49)	(-10.06)
cut4	-0.627***	-0.0607
	(-27.29)	(-0.30)
N	87882	87882
LL	-95882.9	-92964.9
Pseudo R^2	0.112	0.139

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced; L_A, L_D refer to length of delay at arrival and departure, SJT: scheduled journey time, Seat=0 represents a standing passenger

7.3.2. The 'OD' model

Whilst passenger-level models reported in section 7.3.1 were concerned with the impacts of a single on-time or delayed journey on passenger satisfaction, the OD model examines the impacts of average performance on the proportion of satisfied passengers, using the same three versions of satisfaction (i.e. V1-3) as before.

Passenger responses were sorted by origin and destination pair and subsequently aggregated and averaged at the station-to-station level for OD pairs with at least 25 passenger responses. 676 such OD pairs were identified. It is noted that as NRPS is generally representative of the rail trips in the UK, selecting OD pairs with only 25 or more responses means that only the more popular flows are represented in the sample. Moreover, as the responses are averaged across each of the flows, each flow contributes equally to the results. However, the NRPS is not necessarily representative of the delay distribution across the network, thus it is not considered appropriate to account for flow sizes. Nevertheless, it must be highlighted that caution is needed when using the model outputs to generalise about the network-wide impacts. Table 42 shows a correlation matrix of the variables of interest. It can be seen that arrival delay is negatively correlated with the proportion of satisfied passengers using the three definitions of delay satisfaction. Table 43 summarises the variables of interest for the sub-sample of OD pairs retained for analysis and Table 44 shows the distribution of responses per OD pair. Average passenger lateness ranges from 0 to 8 minutes with a mean value of 2.85 minutes. This is similar to both the APL levels reported across the whole NRPS dataset as well as to those recorded at the network level in ORR (2022) APL statistics. Across the OD pairs in the sub-sample, an average journey time of 60 minutes was recorded whilst around 10% of passengers were standing and 90% were seated.

A fractional response logit model was estimated for the three definitions of delay satisfaction (i.e. V1-3). Estimated coefficients are reported in Table 45. The number of journey purpose categories was reduced to just three (business, commute and leisure) to facilitate interpretation and due to relatively small and insignificant differences between journey purpose sub-categories suggested by the passenger models of satisfaction in section 7.3.1. as well as to take into account the effectively reduced sample size. The purpose of the constant varying by journey purpose is to control for differences in how passengers score their satisfaction with punctuality in the case of ‘no delay’. Furthermore, an interaction between lateness and scheduled journey time and an interaction between lateness and the proportion of seated passengers enter the model as explanatory variables.

It is noted that the ‘OD’ models have relatively low pseudo R^2 values when comparing with the ‘passenger’ models. However, these models are expected to perform worse as they use values that are averaged across all the responses within each of the OD pairs, thus ignoring the distribution of the averaged values. It also means that while a general relationship was estimated, there are multiple cases where the predicted proportions of satisfied passengers are much higher or lower than the actual values.

Whilst APL is suggested to have a negative impact on the proportion of satisfied passengers, all else being equal, the impact of journey length is (in most cases) positive, but insignificant. At the same time, the proportion of seated passengers has a significant impact on satisfaction only in the case of commute. The insignificance of some of the coefficients may be a result of the OD model being:

1) based on averages and proportions (thereby moderating the variance in the data) and

2) based on a reduced sample size (i.e. around 700 OD pairs so as to focus on flows with a reasonable number of passenger responses).

Nevertheless, the coefficients are of the expected signs, but only the APL coefficients are statistically significant for all journey purposes.

The proportion of passengers satisfied with punctuality was plotted for the three definitions of delay satisfaction (V1-3) in Figure 49 for a typical (average) journey. Figure 50 on the other hand shows the proportion of passengers satisfied with a given level of performance using Version 2 of delay satisfaction by journey purpose.

Under V2 of delay satisfaction, the proportion of commuters satisfied with performance for a given level of APL is typically lower, in line with the previously estimated models. Whilst around 80% of commuters would typically be satisfied with performance if all trains run on-time, the proportion decreases to 50% for APL of around 5 minutes, suggesting an average decrease of 6 percentage points (pp) per minute of average lateness. For the other journey purposes, the proportion of satisfied passengers is above 90% under perfect performance (i.e. APL=0), but at an APL of around 5 minutes, satisfaction decreases to around 80%, on average by 2 pp per minute of average lateness. This means that the reduction in the proportion of satisfied passengers is also more pronounced for commuters than it is for other travellers.

This would suggest that for any two OD pairs with comparable levels of APL, it is the journey purpose split (i.e. especially the proportion of commuters) that will have the most impact on differences in passenger satisfaction. Therefore, if the proportion of commuters using a given OD flow increases, all else being equal, then levels of satisfaction are likely to reduce. It is also worth highlighting that, whilst an explicit journey length effect could not be clearly discerned, longer journeys will generally be subjected to higher absolute APL (0.4 correlation in the sample) and, at the same time, they will involve a smaller proportion of commuters. Therefore, journey length is likely to have an indirect impact on the proportion of satisfied passengers via related variables.

The models were re-estimated using all levels of the interacted variables (V1_AL-V3_AL). For the V2_AL model, as shown in Annex II (section D), at lower levels of APL, OD pairs with longer journey times are suggested to have a lower proportion of satisfied passengers what changes at higher levels of APL. Similarly to the previously estimated models, this suggests that when all levels of the interacted variables are included, the direction of the estimated relationship is not always plausible, highlighting the problems with using the fully specified models. Due to this reason, the original version of the model has been retained as the preferred version.

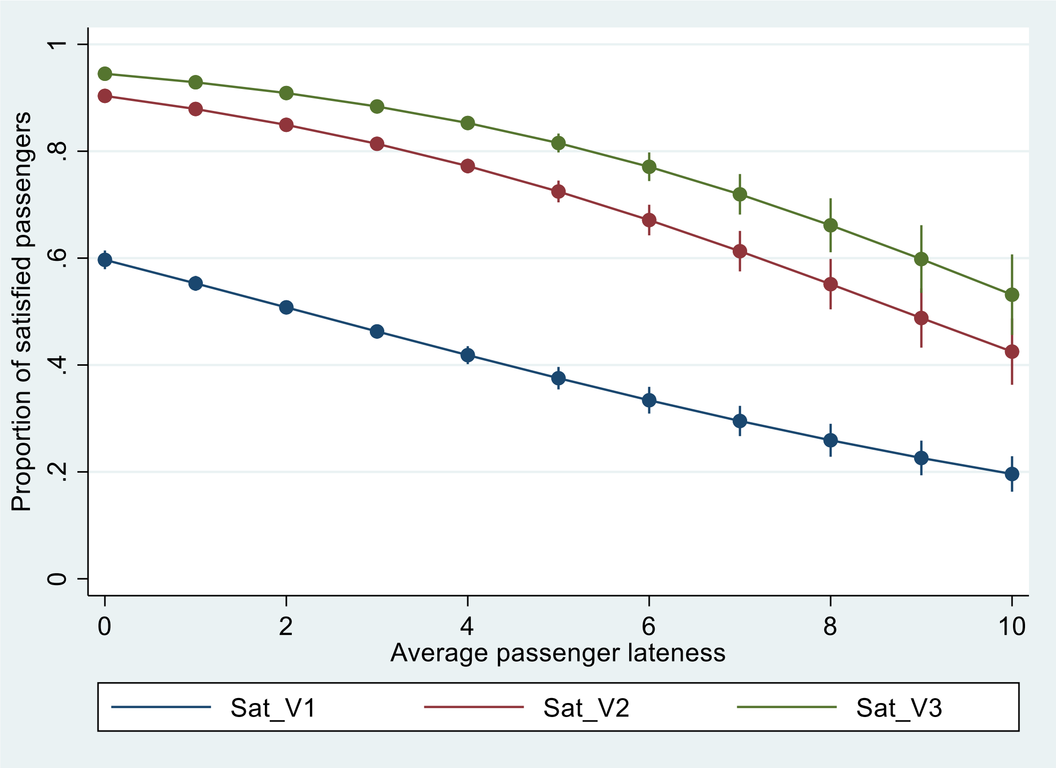


Figure 49 Levels of 'delay satisfaction' for increasing average delay based on the three definitions of delay satisfaction V1: (5) vs (1-4) ; V2: (4-5) vs (1-3); V3: (3-5) vs (1-2) based on the model from Table 45

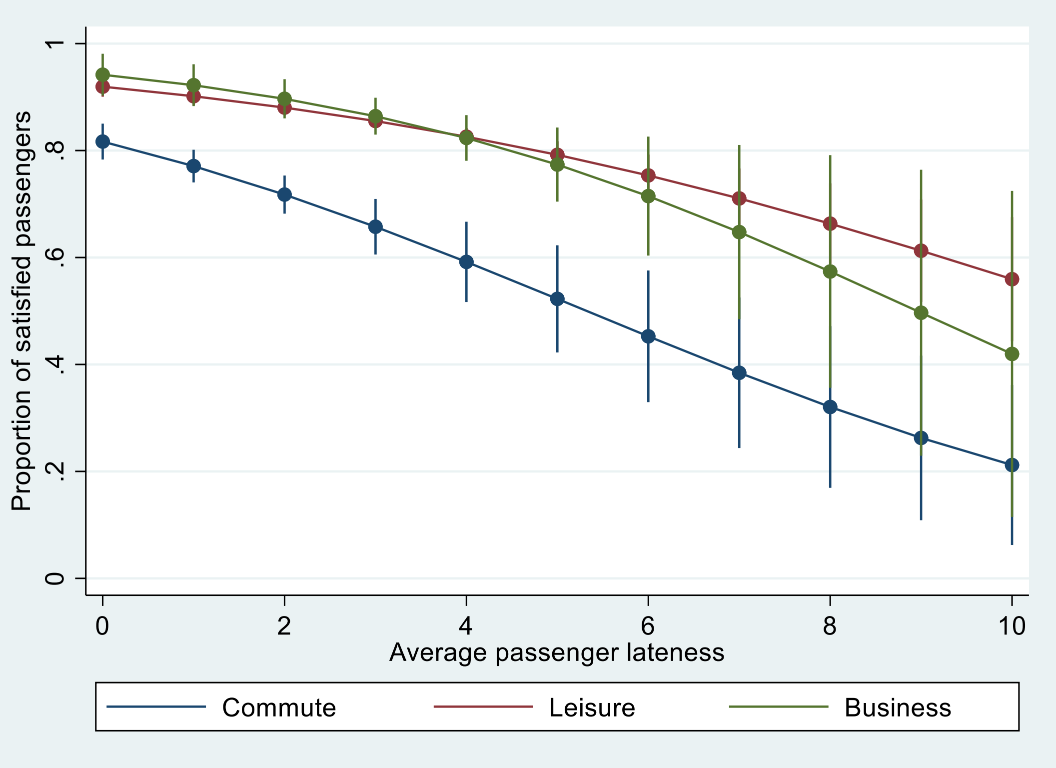


Figure 50 Proportion of satisfied passengers under Version 2 of delay satisfaction at the average values of control variables

Table 42 Correlation matrix

Variables	**(1)**	(2)	**(3)**	**(4)**	**(5)**	(6)	(7)	(8)	(9)	**(10)**	(11)	(12)	**(13)**
(1) APL (arrival)	1.000												
(2) D15	0.802	1.000											
(3) Punct (Sat_V1)	-0.144	-0.021	1.000										
(4) Punct (Sat_V2)	-0.316	-0.160	0.829	1.000									
(5) Punct (Sat_V3)	-0.337	-0.179	0.759	0.933	1.000								
(6) Overall (Sat_V1)	-0.081	0.036	0.737	0.619	0.563	1.000							
(7) Overall (Sat_V2)	-0.160	-0.047	0.719	0.732	0.712	0.716	1.000						
(8) Overall (Sat_V3)	-0.221	-0.112	0.599	0.683	0.702	0.558	0.806	1.000					
(9) APL (departure)	0.605	0.545	-0.391	-0.536	-0.546	-0.142	-0.270	-0.330	1.000				
(10) SJT	0.365	0.338	0.490	0.405	0.376	0.277	0.303	0.254	-0.154	1.000			
(11) PStated	0.681	0.454	-0.597	-0.740	-0.732	-0.380	-0.505	-0.538	0.727	-0.133	1.000		
(12) PRecord	0.768	0.416	-0.276	-0.378	-0.378	-0.200	-0.237	-0.259	0.451	0.157	0.677	1.000	
(13) PSeat	0.016	0.095	0.558	0.570	0.546	0.423	0.520	0.547	-0.285	0.451	-0.393	-0.134	1.000

Legend: APL: average passenger lateness; D15: proportion of delays over 15 minutes; Punct/Overall: punctuality/overall satisfaction; V1-V3: three versions of binary representations of satisfaction; SJT: scheduled journey time; PStated, PRecorded, PSeat: proportion of passengers reporting being delayed, being matched a delay and reporting having a seat respectively; variable names in bold refer to the variables used in models from Table 45.

Table 43 Variable summary (Variable numbers correspond to variables presented in Table 42)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	B	C	L
Mean	2.85	0.04	0.49	0.81	0.88	0.39	0.85	0.94	1.68	61.33	0.21	0.55	0.91	0.19	0.28	0.53
SD	1.48	0.04	0.15	0.12	0.09	0.12	0.09	0.05	1.16	49.72	0.12	0.18	0.10	0.14	0.29	0.23
Min	-	-	0.03	0.39	0.44	0.07	0.48	0.66	-	3.34	-	-	0.35	-	-	-
Max	8.00	0.24	0.85	1.00	1.00	0.78	1.00	1.00	6.48	309.64	0.66	0.97	1.00	0.67	1.00	0.98

Legend: B, C and L: proportion of business, commute and leisure travellers respectively; variables in bold refer to the variables used in models from Table 45.

Table 44 Distribution of the number of responses per OD pair (N>25)

	Mean	SD	Min	Max	P25	P50	P75	N
Responses	83	133	25	1554	32	47	79	676

Table 45 Estimated coefficients of the OD model with control variables

		V1	V1_AL	V2	V2_AL	V3	V3_AL
JP	**Constant**	0.546*	-7.855*	2.783***	-5.753	3.296***	-2.583
		(2.03)	(-2.10)	(7.65)	(-1.44)	(7.40)	(-0.59)
	Commute	-0.886**	7.595*	-1.289***	6.685	-1.141*	4.920
		(-3.02)	(2.02)	(-3.32)	(1.62)	(-2.43)	(1.10)
	Leisure	0.178	8.501	-0.348	5.167	-0.250	1.208
		(0.51)	(1.81)	(-0.75)	(0.97)	(-0.45)	(0.20)
Commute	APL	-0.602***	-0.565**	-0.674***	-0.423*	-0.776***	-0.78***
		(-5.04)	(-2.58)	(-7.48)	(-2.21)	(-7.96)	(-3.48)
	SJT		-0.0109		-0.00952		-0.00243
			(-1.48)		(-1.11)		(-0.24)
	PSeat		0.368		1.225		0.104
			(0.50)		(1.57)		(0.11)
	APLxSJT	-0.00081	0.00257	0.00160	0.00420	0.00189	0.00212
		(-0.70)	(0.99)	(1.31)	(1.45)	(1.41)	(0.63)
	APLxPSeat	0.395**	0.217	0.325**	-0.117	0.392**	0.335
		(2.70)	(0.73)	(2.60)	(-0.41)	(2.92)	(0.99)
Business	APL	-0.639*	0.824	-0.923***	0.291	-1.038***	-0.333
		(-2.01)	(1.19)	(-3.93)	(0.41)	(-4.69)	(-0.44)
	SJT		0.00344		0.00632		0.00504
			(0.55)		(0.65)		(0.46)
	PSeat		8.619*		8.483		5.636
			(2.14)		(1.92)		(1.17)
	APLxSJT	0.00062	-0.00023	0.0025**	0.00071	0.00200	0.00055
		(1.13)	(-0.15)	(2.73)	(0.29)	(1.57)	(0.19)
	APLxPSeat	0.518	-0.978	0.507	-0.634	0.658	0.0701
		(1.42)	(-1.27)	(1.55)	(-0.76)	(1.78)	(0.08)
Leisure	APL	-0.474**	-0.323	-0.441**	0.465	-0.376*	0.861
		(-3.10)	(-1.01)	(-2.75)	(1.17)	(-2.12)	(1.75)
	SJT		0.0001		0.00140		0.00236
			(0.04)		(0.35)		(0.51)
	PSeat		0.0721		3.153		4.660*
			(0.05)		(1.66)		(1.99)
	APLxSJT	0.0007**	0.0007	0.00067	0.00047	0.00118*	0.00080
		(3.16)	(1.13)	(1.93)	(0.50)	(2.23)	(0.72)
	APLxPSeat	0.321	0.163	0.198	-0.761	0.0709	-1.251*

	(1.87)	(0.46)	(1.05)	(-1.74)	(0.33)	(-2.30)
N	676	676	676	676	676	676
LL	-448.1	-447.8	-310.2	-309.5	-236.8	-236.3
Pseudo R^2	0.0433	0.0439	0.0555	0.0576	0.0604	0.0624

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
APL: average passenger lateness; SJT: scheduled journey time; PSeat: proportion of seated passengers; V1: (5) vs (1-4) ; V2: (4-5) vs (1-3); V3: (3-5) vs (1-2); AL refers to models with all interacted variables included

7.4. Summary

NRPS passenger satisfaction data represents a rich dataset, which supports the elicitation of new insights on passenger journey satisfaction, especially in the context of rail performance. All of the alternative approaches to modelling satisfaction introduced as part of this chapter are summarised in Table 46, highlighting the novelties and key insights from each of them.

Table 46 Summary of the estimated models

Model	Approach	Novelty	Key insights
Passenger	**Binary**	Binary representation allows easier interpretation of the satisfaction variable and estimation of the lengths of delay where travellers become more likely to be dissatisfied.	Commuters more likely to be dissatisfied after 5-6 minutes of delay, for other travellers this threshold is between 12-20 minutes Having a seat increases the dissatisfaction threshold by between 2 to 6 minutes for commuters and up to 15 minutes for other travellers.
Passenger	**Ordered (Overall Satisfaction)**	This model is based on a similar analysis by Monsuur et al. (2021), however with a few modifications, most importantly, inclusion of responses where delays were unperceived.	Large differences in the choices between commuters and other travellers with commuters base category being ‘fairly satisfied’ for delays of up to 15 minutes, then changing to dissatisfied Other travellers’ base category is ‘fairly satisfied’ for up to 30 minutes. This is suggested to be earlier than the 50 minutes suggested by Monsuur et al. (2021).
Passenger	**Ordered (Punctuality Satisfaction)**	Using a variable directly related to satisfaction with punctuality, not the whole journey, allows studying the indirect impacts of journey quality and/or	Punctuality satisfaction is typically lower than overall satisfaction. ‘Satisfied’ categories are dominant for business and leisure travellers for delays of up to

		length on how delays impact passengers rather than their impacts on overall satisfaction.	20-25 minutes. For commuters, ‘dissatisfied’ categories already become dominant for delays of over 10 minutes.
OD	**Binary**	The binary representation at the aggregated OD levels allows estimating the impact of average performance on the proportion of satisfied passengers for a given OD pair.	Under perfect performance, 80% of commuters and 90% of other travellers are suggested to be satisfied with performance. The proportion of satisfied passengers decreases by (on average) 6 pp for a minute of APL for commuters and 2 pp for other travellers. Hence, commuter-focused OD pairs are likely to suffer from lower levels of passenger satisfaction given the same level of APL.

In summary, the binary model of punctuality satisfaction under Version 2 of delay satisfaction is the recommended model with the OD model being possibly more useful for policy applications as it can be used for benchmarking and forecasting the impacts of performance on the proportion of satisfied passengers. The estimated thresholds could be used for setting performance targets where the ‘passenger’ models can be useful in suggesting the optimal, yet attainable, distribution of delay incidents of given lengths (i.e. delays of more than 15 or 20 minutes are more likely to lead to choices of the categories related to lower satisfaction). At the same time, the ‘OD’ model may be helpful in setting average performance targets, more in line with the typical focus of the industry. Finally, both models can be applied to set the performance targets, both in terms of the average performance and the distribution of delay length occurrences. Nevertheless, it must be noted that aggregating data (as in the case of the ‘OD’ model) as well as dichotomising the data (as in the case of the binary version of the ‘passenger’ model) lead to some loss of detail and especially so related to the strength of satisfaction versus dissatisfaction.

In terms of the reported satisfaction, commuters are found to be considerably more sensitive to delays, being unlikely to express complete satisfaction even in the absence of delay. Therefore, station-to-station journeys with higher proportions of commuters are likely to be associated with lower levels of passenger satisfaction with delays. Whilst commuters express most dissatisfaction with performance, they are least responsive to performance in demand/revenue terms. This is likely because of a lack of viable travel alternatives for commuters, which renders them captive to rail (Batley et al., 2011).

Crowding (expressed as the ability to find a seat) is found to be an important confounding factor, such that dissatisfaction with delay will be compounded if passenger is stood rather than seated. Interestingly, the impact of scheduled journey time for a given delay length is suggested to be positive on delay satisfaction for seated business and leisure travellers. In the case of standing passengers, this impact is typically smaller. This may be due to the ability to use travel time more productively when seated (as discussed by Wardman and Lyons, 2016; Lyons et al., 2016) as well as larger safety buffers around arrival times or lower sensitivities to delays related to the type of activity planned. For commuters, however, the impact of scheduled journey time for a given delay length is less clear and, in one case, even negative (though insignificant). This may be connected to the suggestions made by Cats et al. (2015) that long commute is generally associated with lower satisfaction with public transport.

The present work focused on understanding how different levels of incidental lateness and average performance affect passenger satisfaction. One area for future research could be to explore the scope to more explicitly link performance to satisfaction and demand. The main limitation of the NRPS dataset in this context is, however, its cross-sectional nature. Each NRPS record represents a given passenger's satisfaction with a given incidence of delay. In the OD level model, this means that the proportion of satisfied passengers does not represent passenger satisfaction with average performance, but rather average satisfaction having averaged across all of the lateness incidents encountered by passengers. This property of the data does not readily lend itself to reconciliation with other (more established) performance metrics, such as AML, proportion of stops delayed by a given amount of minutes or similar supply-centric measures. This issue could potentially be addressed in the future, by collecting satisfaction data from a panel of commuters over a period of time to better understand the relationship between average satisfaction, incidental satisfaction and how both are affected by average delay length and its distribution.

As previously discussed, another important limitation of NRPS also lies within its inability to represent cancellations or interchanges. The impact of these data errors was investigated more closely in the case of the analysis of delay perception. However, the results from sensitivity analyses were broadly in line with the main body of analysis.

It has been suggested that commuters tend to be less satisfied with performance. This could be due to strategic bias or their reflection of general performance on the specific OD pair – being especially evident in relatively lower satisfaction levels for the 'no delay' case, as well as a more rapid decrease in satisfaction levels as delays increase. These lower satisfaction levels do not, however, appear to stimulate a direct response in terms of demand. Further work is required to better understand this conundrum and how it can be represented in social welfare terms – so as to focus public investment in the railways where it will achieve best value for money.

The satisfaction analysis presented useful insights into the levels of delays that are detrimental to passenger satisfaction for different types of passengers as well as the impacts of journey lengths and quality on the disutility of delay. Following the separate analysis of delay perception and satisfaction using comparable methodologies, the next step is to ask about the differences between the probabilities of perceiving a delay and then being dissatisfied with it – i.e. with an a priori expectation of a gap between the moment the delay is perceived and when it starts affecting passenger satisfaction. This analysis will be conducted as part of the next section.

7.5. The gap between delay perception and dissatisfaction

The main focus of this chapter is on comparing the concepts of delay perception and satisfaction by investigating how the predicted probabilities of delay perception and journey satisfaction change for increasing lengths of delays. This will be based on a comparison of the results from perception and satisfaction models presented in Chapter 6 with the passenger model of delay satisfaction using V2 of the satisfaction variable (i.e. 'very satisfied' or 'fairly satisfied' versus other options). The comparison will be based on the extended versions of the perception and satisfaction models that use the same control variables, allowing direct comparisons between them. The predicted delay length perception and satisfaction thresholds are subsequently compared where the gap between delay perception and its impacts on satisfaction is defined as the difference between the length of delay with an estimated 0.5 probability of perceiving and being satisfied with it.

Considering the relative complexity of the logistic regression models with multiple explanatory continuous variables and their interactions as well as a relatively large number of journey purpose categories used in the previously estimated models, an attempt was also made to estimate a simplified version of the model of perception and satisfaction that would facilitate comparisons between the two. The simplified models were estimated with:

1) Only three journey purpose categories – business, commute and leisure;

2) Reduced number of explanatory variables to facilitate interpretation. The binary outcome referring to perception or satisfaction is now only explained by an interaction between the length of delay at arrival and a dummy variable

representing whether passenger was standing or seated, allowing for heterogeneity by journey purpose;

3) Satisfaction variable being recoded with the outcome taking the value of 1 if a passenger was dissatisfied or 0 if a passenger was satisfied (previous satisfaction models used the opposite representation) to better align with the perception modelling framework where the probabilities increase with experienced delay lengths. This represents a rather cosmetic difference, facilitating graphical comparisons but having no impact on the econometric results.

Hence, the purpose of this section is to uncover the intermediate stages of the impacts of delays on passengers by:

1) establishing links between delay occurrence, perception and satisfaction to enable better understanding of these concepts and
2) provide an estimate of a gap between delay perception and its impacts on satisfaction that is defined as the difference between the length of delay that is perceivable and the corresponding delay that is deemed detrimental to passenger satisfaction.

This section is divided into a subsection reporting the results of the comparison between perception and satisfaction models (7.5.1) and a summary of results (7.5.2).

7.5.1. Results

To compare the concepts of delay perception and satisfaction, the model results are used to present:

1) the threshold of delay perception, i.e. where the estimated probability of perceiving a delay reaches 0.5,
2) the threshold of delay dissatisfaction, i.e. where the estimated probability of being dissatisfied with punctuality increases to 0.5 and
3) the corresponding difference between the two thresholds that represents the gap between delay perception and dissatisfaction

This is first introduced by comparing thresholds based on the previously estimated models and subsequently by estimating simplified versions of these models to facilitate reporting and interpretation.

Comparison of the extended models introduced in Chapter 6 and section 7.3

Binary models of delay perception and satisfaction (using V2 of the binary variable) in the extended version were previously estimated and reported in Table 26 and Table 36.

Based on the estimated models, the probabilities of perceiving a delay and being satisfied with it were predicted for:

- six journey purposes,
- 10^{th} and 90^{th} percentile of scheduled journey time distribution (for each of the segments),
- 0 and 15 minutes of departure delay,
- seated and standing passengers and
- arrival delay length of 0-30 minutes.

Subsequently, the arrival delay lengths were evaluated at 0.5 probability of perception and satisfaction and are reported in **Figure** 51. By comparing the two thresholds (i.e. of perception and satisfaction), this figure highlights the gap between delay perception and satisfaction, as indicated by the error bars. As expected, in most cases, the predicted delay length thresholds of satisfaction are larger than those of perception. This would be indicative of a gap between the moment when a delay is perceived and when it starts having a detrimental effect on passenger satisfaction. In the cases where the gap is suggested to be negative, this is due to the perception and satisfaction thresholds not being significantly different from each other. Moreover, as the perception variable truly represents a binary outcome, satisfaction variable is dichotomised and, as previously noted, there is more than one way to convert an ordinal scale into a binary one. This highlights the relative complexity of the extended versions of the models due to their multidimensional nature. However, it might be worth highlighting some key points resulting from the analysis:

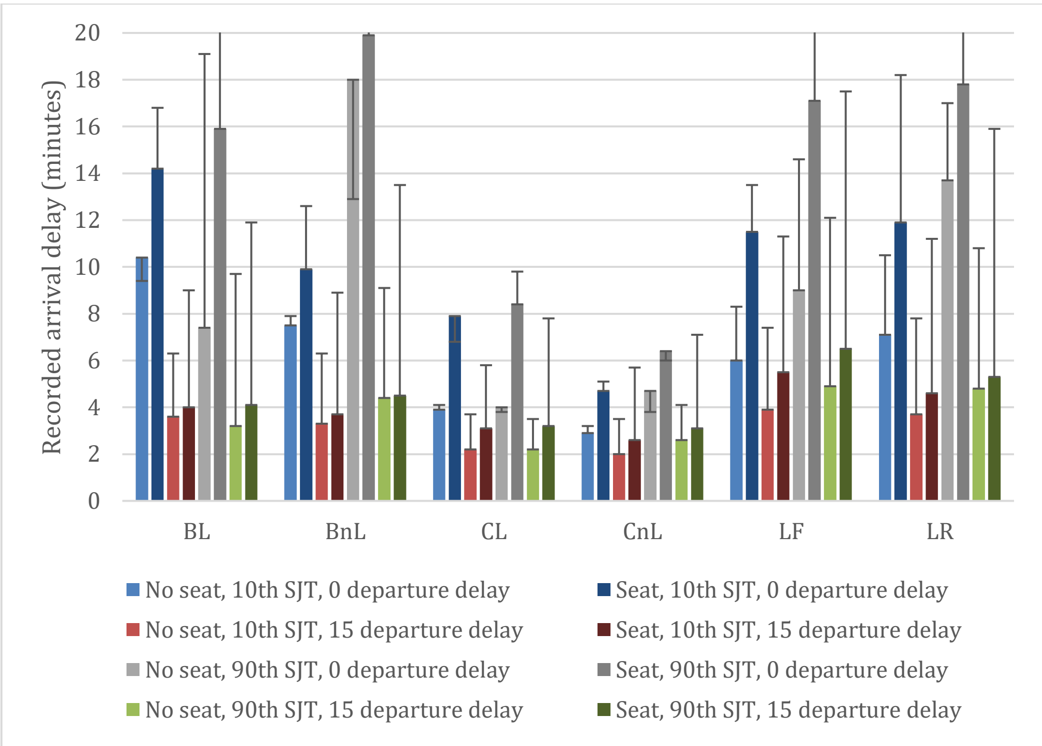
1) Perception and satisfaction thresholds for commuters are typically lower than 10 minutes for all the studied scenarios.
2) The respective thresholds for other travellers are typically between 4 and 20 minutes. However, for long journeys with no delay at departure, these might be in excess of 20 minutes if passengers had a seat.
3) Delay and perception thresholds are typically insignificantly different for very short journeys and/or in the cases where the train departed on time from the origin.

This would suggest the important role of departure delay that affects delay perception more than satisfaction.

4) The gap between delay perception and satisfaction is relatively small for commuters, from less than 1 minute in the case of journeys with no delay at departure to around 4-5 minutes for longer journeys with a longer delay at departure and for seated passengers.

5) For other travellers, the gap is typically small for short journeys with no delay at departure and for standing passengers (i.e. 0-2 minutes), but typically increases for seated passengers, with delay at departure (at least for short journeys as there is less confidence in the estimates for longer journeys) as well as with journey lengths.

The next section aims to provide a clearer picture by introducing simplified versions of delay perception and satisfaction models that facilitate interpretation.



Legend: perception thresholds indicated by column heights whilst error bars demonstrate the difference between perception and dissatisfaction thresholds. The example interpretation of the figure presented above would be that (for example, second column) for Business travellers to London with a seat, no delay at departure and journey length equal to the 10th percentile of journey length distribution, the probability of perceiving a delay is equal to 0.5 when a 14-minute delay is experienced. The corresponding delay length leading to the probability of dissatisfaction being equal to 0.5 is just under 17 minutes, hence, the difference between the two delay length thresholds is around 3 minutes as shown by the error bar.

Figure 51 Estimated delay length perception and satisfaction thresholds

Estimation of the simplified models of delay perception and satisfaction

Perception and satisfaction models in their extended form are relatively complex. This is primarily due to a large number of continuous variables and their interactions. To mitigate this, simplified models of delay perception and satisfaction were also estimated for only three journey purposes and two explanatory variables (interaction between a dummy variable representing whether a passenger was seated or standing and the length of delay at arrival). The reported coefficients are presented in Table 47 with the predicted probabilities depicted graphically in Figure 52. The pseudo R^2 values are generally of a magnitude that is comparable to the models reported throughout the thesis. In both cases, the estimated coefficients are of expected signs and magnitudes.

As was the case with the previously estimated models, the predicted delay perception and dissatisfaction thresholds are reported (Table 48). These are comparable for business and leisure travellers. As such, standing business and leisure travellers become more likely to perceive delays of around 6 minutes as compared to just under 10 minutes for seated passengers. The estimated gap between delay perception and dissatisfaction is around 2.5-4 minutes for standing passengers and 7.5-9 minutes for seated passengers. This suggests that journey quality not only has an impact on delay perception and satisfaction separately – having a seat also increases the difference between the two. This is also in line with the previously reported results. Commuters are typically able to perceive smaller delays – around 3-5 minutes with the estimated gap between delay perception and dissatisfaction being much smaller, insignificant for standing passengers and just above 1 minute for seated passengers. This highlights the smaller impact of journey quality on commuters as well as these travellers being generally much more sensitive to delays. Moreover, the small gap is indicative of almost all the perceivable delays automatically affecting commuters' satisfaction.

The models were also re-estimated using all levels of the interacted variables (models Perc_AL and DSat_AL). The main difference between the originally estimated models and the fully specified models is in the estimated relationship between seat availability and the probability of perceiving a delay, and being dissatisfied with it. When all levels of interacted variables are included, this allows for the base levels to be different for seated and standing passengers. However, this also means that the model does not only capture the complementary nature of being seated versus standing on how delay is perceived or affects satisfaction as the baseline probabilities are assigned individually for both levels of the dummy variable.

Table 47 Estimated binary model coefficients for perception and dissatisfaction models

		Perc	Perc_AL	DSat	DSat_AL
	Constant	-1.540***	-0.945***	-2.047***	-1.022***
		(-47.17)	(-7.28)	(-56.80)	(-8.21)
Journey purpose	Commute	0.656***	0.561***	1.382***	0.978***
		(15.13)	(3.92)	(30.90)	(7.16)
	Leisure	-0.148***	0.125	-0.229***	-0.145
		(-3.71)	(0.80)	(-5.14)	(-0.95)
	Seat=1				
	Business		-0.630***		-1.099***
			(-4.69)		(-8.45)
	Commute		-0.631***		-0.782***
			(-9.18)		(-12.22)
	Leisure		-0.922***		-1.197***
			(-10.12)		(-13.19)
	Arrival delay				
Business	(Seat=0)	0.236***	0.160***	0.226***	0.116***
		(14.92)	(7.76)	(15.61)	(6.87)
	(Seat=1)	0.163***	0.167***	0.121***	0.127***
		(34.94)	(34.91)	(29.70)	(30.34)
Commute	(Seat=0)	0.294***	0.204***	0.227***	0.130***
		(26.93)	(15.39)	(23.77)	(12.01)
	(Seat=1)	0.176***	0.195***	0.104***	0.124***
		(30.22)	(30.34)	(22.24)	(24.00)
Leisure	(Seat=0)	0.277***	0.162***	0.225***	0.114***
		(24.16)	(11.48)	(23.98)	(10.31)
	(Seat=1)	0.172***	0.177***	0.120***	0.127***
		(52.17)	(52.36)	(43.17)	(44.21)
	N	48904	48904	48904	48904
	LL	-28062.1	-27965.0	-24617.7	-24438.7
	Pseudo R^2	0.136	0.139	0.145	0.151

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
Seat=0 represents a standing passenger; AL refers to the model with all levels of interacted variables; Perc refers to perception and DSat - dissatisfaction

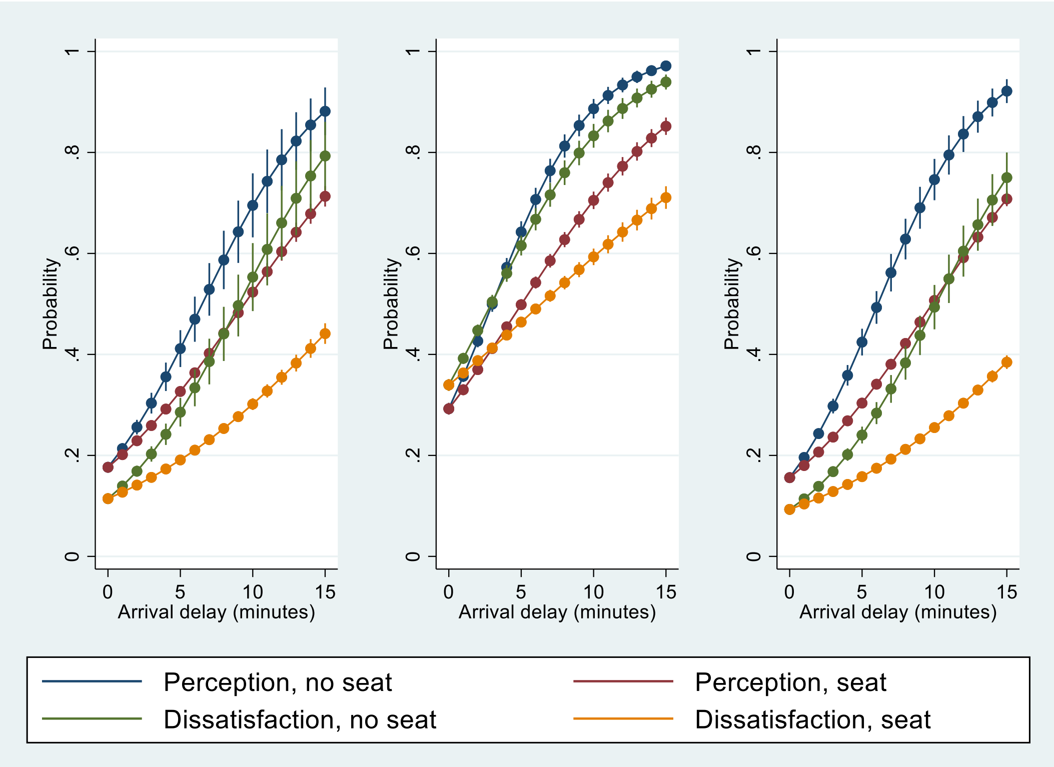


Figure 52 Probability of delay perception and dissatisfaction for increasing delay lengths and different journey purposes based on the models presented in Table 47

Table 48 Estimated perception and dissatisfaction delay length thresholds based on models presented in Table 47 and Figure 52

	Perception		Dissatisfaction		Gap	
Journey purpose	**No seat**	**Seat**	**No seat**	**Seat**	**No seat**	**Seat**
Business	6.5	9.4	9.0	16.9	2.5	7.5
Commute	3.0	5.0	2.9	6.3	-0.1	1.3
Leisure	6.1	9.8	10.1	18.9	4.0	9.1

7.5.2. Summary

The previous section compared two concepts introduced in the thesis, i.e. delay perception and satisfaction, introducing the concept of a gap between delay perception and dissatisfaction. The predicted probabilities of perceiving a delay and being dissatisfied with it were compared to investigate the existence of a gap between:

1) the moment of delay occurrence,

2) the delay lengths that start being perceivable by travellers and

3) the delay lengths that start having a negative impact on passenger satisfaction.

The so-called gap between delay perception and satisfaction was first compared for a more complicated model with several control variables and a simpler model for easier interpretation of the results.

The analysis suggests existence of a gap between the moment when delays are perceived and start having impacts on passenger satisfaction. The gap is suggested to be between 0 minutes for standing commuters to up to 9 minutes for seated leisure travellers. Typically, commuters are able to perceive smaller delays than other travellers, with smaller delays also having a more significant impact on their satisfaction levels. Both the thresholds of perception and dissatisfaction as well as the estimated gap between them are very similar for business and leisure travellers. It has been generally suggested that the gap between delay perception and dissatisfaction increases for seated passengers, with longer journeys and delay at departure.

There are some limitations to the approach undertaken as part of this chapter:

1) Most emphasis is placed on the thresholds where probabilities of delay perception and dissatisfaction reach 0.5. This may be a useful approach due to its simplicity, however, it can be seen that probabilities change marginally for each incremental increase in delay length. There might be some value in understanding how the so-called gap changes with different lengths of recorded delays.

2) Only one version of the binary representation of the dependent variable was used. This was based on the analysis conducted in the previous chapter as well as practical considerations. However, the previous analysis detailed and commented on the alternative approaches to the treatment of the reported satisfaction variable.

The differences between the moments when delays are perceived and start having significant impacts on passenger satisfaction call for a more detailed analysis of the marginal impacts of delays on passengers. If very small delays are unperceived and there is a tolerance threshold for delays before passengers become dissatisfied, this would possibly suggest that the marginal impacts of smaller versus larger delays may be, in fact, different. As such, a natural recommendation that arises from this analysis is to study the possible non-linearities in the marginal utility of late time as discussed in the next section.

7.6. Marginal (dis)utility of lateness

Previous sections focused on establishing a link between delay length and passenger satisfaction with the use of logistic regression. The concepts of delay perception and satisfaction were compared, suggesting that smaller delays are often likely to be unperceived and noting that perceiving a delay does not automatically lead to a consequential impact on passenger satisfaction. This, in turn, leads to posing a question about the curvature of the shape of the relationship between delay and travellers' utility.

Before introducing the possible ways of investigating the non-linearities in the impact of delays on utility, it is worth highlighting that most of the currently used methodologies in economic appraisal assume constant valuations of time and delays (e.g. value of time or reliability multipliers later introduced in Chapter 8). However, some indications of non-linear unit valuations of late time were suggested in the literature, e.g. by Wardman and Batley (2022), arguing that proportional elasticities (i.e. based on the relative proportion of AML to GJT) better explain changes in demand than the actual delay lengths.

Hence, the purpose of this section is to explore the potential non-linearities in the delay impacts using approaches introduced by previous literature. In doing so, it needs to be emphasised that such analysis is to be conducted with data on reported satisfaction used as a proxy for utility. One important characteristic of the satisfaction data that needs to be accounted for is its ordinal nature. Throughout the literature, different methodologies have been applied to modelling ordinal data from satisfaction surveys. Gao et al. (2018) summarised years of previous research concerning journey satisfaction, observing that most of the studies are empirically-driven where the choice of the functional relationship is made at the discretion of researchers. Hence, there is an abundance of studies modelling satisfaction using both the original ordinal scales (i.e. Cats et al., 2015; Yang et al., 2015; Ettema et al., 2016; Abenoza et al., 2017) or assuming an interval scale and applying linear regression methods (i.e. Cao and Ettema, 2014; De Vos et al., 2016; Wan et al., 2016).

Conceptually, logistic regression methods are more appropriate for modelling the relationship between travel attributes and choice of satisfaction scores. That said, it has been noted that there are several benefits of applying linear regression with the most obvious being the ease of interpretation. Several studies discussed the pros and cons of imposing an assumption of cardinality on ordinal data, the usefulness of such approaches as well as how the results compare between these types of approaches (e.g. Dickerson et al., 2014). In such cases, there is implicitly a very strong assumption of equal distances between the different points on the ordinal scale. There is no consensus in the literature regarding this problem. However, some studies have shown similar results obtained from logistic and linear regressions (e.g. Ferrer-i-Carbonell and Frijters, 2004). This could possibly suggest that it may be worth applying both approaches simultaneously.

The concept of utility dates back to Bernoulli's (1954) work on risk aversion suggesting that while prices for any two individuals are equal, the utility derived by each individual from buying the same good may differ. Similarly, for an individual on a lower income, a same-value gain (income) may be usually more significant than for a person with a higher income. While utility is assumed to be additive, it is important to understand the rate of change in utility levels in response to changes in the variable of interest. This has been studied in various contexts in public policy, usually focusing on the marginal utility of income (Layard et al., 2008) or consumption (Evans, 2005). These estimates are important parameters in determining optimal taxation (Layard et al., 2008) including carbon tax and climate change policy (Anthoff et al., 2009; Bachmann, 2020), assessment of health technologies and valuing health (Phelps, 2019), analysis of labour market (Farzin, 2009; Rätzel, 2012; Masuda et al., 2021), progressive pricing (Coker and Izaret, 2021) and valuation and cost-benefit analysis of infrastructure investments (Greene et al., 2020), including valuation of time in transport contexts (Batley et al., 2019).

The work conducted as part of this section aims to look at how the marginal utility of delay changes with increasing delays. This is done by following methodologies introduced by previous literature, described in more detail in section 7.6.1. Initial analysis of the relationship between delay and reported satisfaction based on the NRPS survey is presented in section 7.6.2 with the results and conclusions of the analysis presented in the remaining sections.

7.6.1. Literature review

When studying the non-linearities in the impacts of delays on passengers, it is worth looking at the methodologies used to study such non-linearities in other areas of economics (e.g. the previously mentioned labour economics). In the case of income, additional earnings can often only be obtained by increasing working hours. As shown by Rätzel (2012), marginal utility of labour follows an inverse U-shape relationship, suggesting that initially, work can increase utility. However, at some point, additional earnings are no longer able to compensate for the increased working hours. Similarly as in the case of labour supply, in transport passengers are concerned with the time and money aspect of travel. Increasing money cost and travel time sources of disutility.

Layard et al. (2008) studied the relationship between household income and life satisfaction to estimate the elasticity of marginal utility of income and understand the marginal impact of income on life satisfaction. The obtained estimates of the elasticity of marginal utility of income of around -1.2 indicated that the marginal utility of income diminishes, contrary to the assumption made by Bernoulli (1954) of marginal utility of income being inversely proportional to income. While subjective well-being or stated life satisfaction is often considered to be a good approximation of utility levels (Layard et al., 2008), Cooper (2020) noted that this approach may discriminate against people with lower expectations or higher willingness to adapt. This reflection in the context of journey satisfaction may be translated to frequent travellers having better experience and knowledge about the typical journey times and average delays or higher sensitivity to lateness. Similarly, the findings of Boyce and Wood (2011) indicate a strong personality effect on the marginal utility of income, suggesting that there might be some levels of heterogeneity in the impacts of additional income on life satisfaction that can only be attributed to respondents' personalities. However, these concerns can be addressed by segmenting travellers. Moreover, subjective well-being or satisfaction largely relies on the assumption of individual rationality which, as noted by Cooper (2020), cannot be guaranteed. On the other hand, subjective well-being has been found to correlate with socio-demographic characteristics, income, health, environment, geography, societal norms and culture (Layard, 2006; Layard et al., 2008; Cooper, 2020; Masuda et al., 2021). Thus, suggesting that the subjective well-being changes in line with the changes in objective measures and highlighting the usefulness of using satisfaction data in economic analysis.

In public transport, the departure times are pre-defined and discrete rather than continuous as is the case with car travel. As discussed as part of section 2.2, scheduling models can be particularly useful in analysing how passengers schedule their journeys and how they are impacted by delays. A choice of departure time depends on the preferred arrival time (PAT) (timing constraints), public transport schedule as well as the safety margin. The size of the safety margin depends on individual preferences, risk aversion and expectations that can depend on previous experiences. Ultimately, passengers aim to maximise expected utility with respect to preferred arrival time but are unlikely to calculate probabilities for all possible travel options, as this is not feasible. The standard formulation used in UK rail assumes that the value of early arrival is 0 and PAT is equal to scheduled arrival (Bates et al., 2001) with the marginal (dis)utility of late arrival being constant. As indicated by Bates et al. (2001), only 1 in 3 travellers has a preferred arrival time equal to scheduled arrival as travellers are likely to include safety margins to their schedules. This could, perhaps, mean that some of the smaller delays are of lower importance as long as a traveller arrives to their destination before the other planned activities start.

Drawing comparisons to income, it is important to note that income is a source of positive utility while additional travel time related to delays is expected to be a source of disutility. Of note, the distribution of delays is also much more skewed to the left with most delays being very small. The reasons for believing that the marginal utility of delay is non-monotonous can be sought in:

1) the analysis of delay perception conducted as part of Chapter 6, suggesting that some smaller delays are unperceived and

2) Wardman and Batley (2022) concluding that elasticities of AML relative to GJT better explain changes in demand than the ones based on the absolute values.

As noted by Gao et al. (2018) different methods have been applied to modelling passenger satisfaction, often chosen at researchers' discretion (typically logistic and linear regression methods). However, Gao et al. (2018) also proposed a cubic model of passenger satisfaction where the explanatory variables were related to differences between experienced and expected travel attributes (related to time components of GJT) for bus users in the Chinese city of Xi'an. Cubic relationships allow studying the different inflection points with marginal utility being non-monotonous.

To decide on the best functional form for studying the relationship between delay and reported satisfaction whilst also examining the potential non-linearities, the next section aims to comment on this relationship using the data from the NRPS survey.

7.6.2. Initial analysis

If one assumes equal intervals between the responses on the original 5-point Likert scale forming part of the NRPS questionnaire then this allows for averaging satisfaction scores across the responses. While there might be some disadvantages to the introduced approach, it is noted that such methods have been used throughout the literature (i.e. in Layard et al., 2008). Moreover, the binary representation of the satisfaction variables, introduced earlier in this chapter, also imposed a strong assumption as multiple scores were grouped, in fact, ignoring the distances on the scale between them. Figure 53 and Figure 54 show the relationship between delay length and reported satisfaction with punctuality and overall journey from the NRPS. The responses were sorted by three journey purposes and delay length, and subsequently averaged across groups of 100s.

As expected, experienced utility (proxied by reported satisfaction) decreases with increasing delay levels with the base satisfaction being lower for commuters whose marginal reduction in utility is also larger (in absolute terms) as discussed earlier in Chapter 7. The utility levels seem to drop in a linear manner. However, towards lengthier delays, the changes become less pronounced.

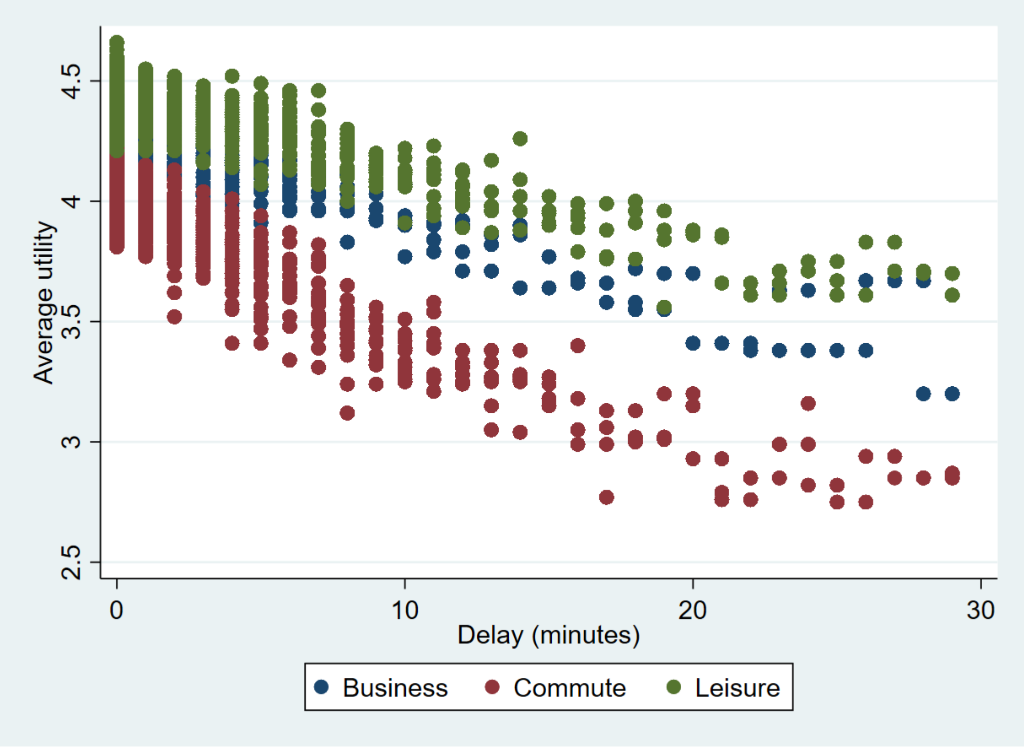


Figure 53 Relationship between delay length and average satisfaction with punctuality (utility). Responses were sorted by journey purpose and delay lengths, and subsequently averaged across responses in the groups of 100 responses.

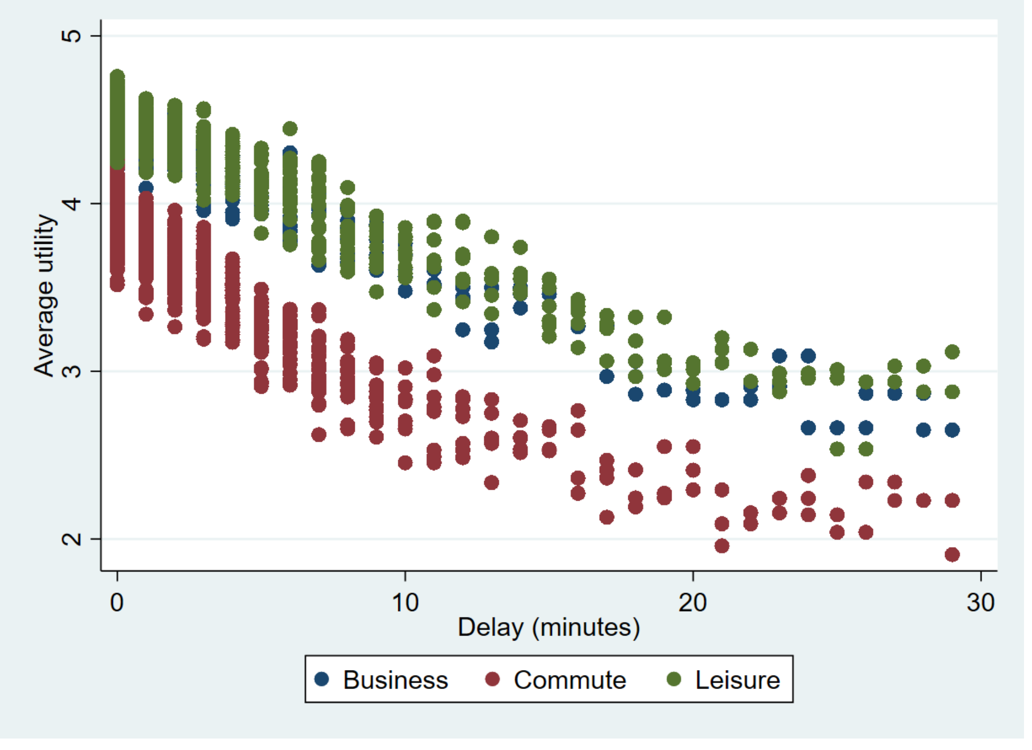


Figure 54 Relationship between delay length and average overall satisfaction. Responses were sorted by journey purpose and delay lengths, and subsequently averaged across responses in the groups of 100 responses.

The methodology section introduces the modelling methods based on the previous literature, also taking into account the observed relationships described above.

7.6.3. Methodology

The key aspect of the investigation conducted as part of this section is to analyse the shape of the relationship between delay and utility of lateness. The first step is to choose a proper functional form that satisfies all the properties of the reported satisfaction data used as a proxy for utility. There is very limited literature covering the possible non-linearities in the delay impacts on passengers - despite Wardman and Batley (2022) considering this issue to be of great importance. On the other hand, there is an abundance of economic literature exploring the impacts of different variables on utility (satisfaction) in a variety of contexts.

The approach undertaken here is based on previous literature, primarily on two studies:

1) Layard et al. (2008) looking at the marginal utility of income and
2) Gao et al. (2018) examining the non-linear relationship between the difference in observed versus expected trip attributes and reported satisfaction.

It is, however, important to consider some of the following properties of NRPS satisfaction data and assumptions, which are additional to the limitations highlighted previously:

1) The utility is based on a difference between experienced and scheduled journey length. This is slightly different to the formulation used in Gao et al. (2018) where a difference between expected and experienced in-vehicle times was used. In the analysis which follows, it is assumed that the expected arrival is equal to the scheduled arrival (as per the timetable). As such, it is expected that any positive difference between actual and scheduled arrival is a source of disutility. It is worth highlighting that in the case of Gao et al. (2018), surveyed respondents were metro or urban bus users with typically higher service frequencies as compared to rail passengers surveyed as part of NRPS.
2) Reported travel satisfaction – i.e. both overall satisfaction and satisfaction with punctuality are proxies for individual welfare. This assumption has generally been used in the literature, especially in the context of subjective well-being (e.g. Kahneman and Krueger, 2006; Layard et al., 2008; MacKerron, 2012).
3) For linear models, it is assumed that 1-point changes in reported satisfaction imply the same distances between the points on the scale. Such assumptions have been also made in the literature (i.e. Ferrer-i-Carbonell and Frijters, 2004; Layard et al., 2008; Dickerson et al., 2014), suggesting that often similar results are obtained

from linear and ordered logit models. However, this approach has also been criticised (e.g. by Baetschmann et al., 2015), hence both cardinal and ordinal representations of the satisfaction (utility) variable are analysed below.

Considering the relationships described previously, the data characteristics and the previous research, the approaches include:

1) Maximum likelihood estimation of ρ using ordered logit and linear dependence model based on the Layard et al. (2008) investigation of the marginal utility of income, i.e.

$$U = \beta_0 + \beta_1 \frac{{L_A}^{1-\rho} - 1}{1 - \rho} + \sum_{n=1}^{n} \beta_n Sat_n$$

(26)

where:

U: experienced utility (i.e. proxied by reported satisfaction, related to overall satisfaction or satisfaction with punctuality);

L_A: recorded length of delay at arrival (destination station)

Sat_n: relates to additional controls, denoting satisfaction with train and station for the models based on overall satisfaction. These are not included in the case of punctuality satisfaction that is related to delay-specific satisfaction (as discussed earlier in the chapter). In the case of the models of overall satisfaction where satisfaction with other journey aspects is used as explanatory variable, it is important to acknowledge the potential endogeneity bias. This is related to needing to control the overall journey satisfaction for those specific journey aspects that are difficult to measure and/or there is no alternative variable that could control for their effect on the overall satisfaction. This potential bias is only a concern for the extended version of the overall satisfaction models.

ρ: minus elasticity of marginal utility of delay

2) Estimation of a cubic model based on Gao et al. (2018) where it is assumed that utility depends on the difference between experienced and expected journey time with experienced journey time being calculated as the difference between actual and scheduled arrival and expected journey time assumed to be as per timetable. Here, a third-degree polynomial regression model is estimated in line with Gao et al. (2018), i.e.

$$U = \beta_0 + \beta_1 L_A + \beta_1 {L_A}^2 + \beta_1 {L_A}^3 + \sum_{n=1}^{n} \beta_n Sat_n$$

(27)

3) Piecewise regression that allows estimating two linear relationships with a breakpoint C. The benefit of fitting such regression is that the breakpoint is not an input of the estimation. This approach is somewhat similar to assuming that at some level of delay, there is a threshold where the marginal impacts of delay start exhibiting a different relationship. The breakpoint is estimated in the model and a second regression model is estimated after the breakpoint C, i.e.

$$U = \begin{Bmatrix} \beta_0 + \beta_1 L_A & for\ L_A < C \\ \beta_0 + \beta_2 L_A & for\ L_A \geq C \end{Bmatrix}$$

(28)

7.6.4. Results

This section presents the estimation results for the models introduced in the methodology section in the following order:

1) Maximum likelihood estimation of ρ
2) Cubic relationship between delay length and satisfaction
3) Piecewise regression

All the aforementioned models aim to investigate the non-linearities in the relationship between delays and utility.

Maximum likelihood estimation of ρ

Following the work conducted by Layard et al. (2008), maximum likelihood estimation of ρ was the first step of the analysis, focusing on analysing the curvature of the relationship between delay and satisfaction (utility). The analysis of the relationship between recorded delay length and average reported satisfaction suggested that the marginal change in satisfaction becomes less pronounced at lengthier delays as well as more noise is visible given that the prevalence of longer delays is more limited. The maximum likelihood estimation has been conducted for the whole sample used throughout the thesis (i.e. delays of up to 30 minutes) as well as for a more restricted version of the sample focusing on delays of up to 15 minutes (chosen as a potential arbitrary threshold based on the graphical analysis of the relationship) using the Apollo package in R (Hess & Palma, 2019). A very pronounced difference in the results from the estimation based on the ‘full’ and ‘restricted’ samples would mean that the studied relationship is much more complex. In the case of income, as studied by Layard et al. (2008), the sample used in the initial version of the modelling had already been restricted by removing the observations in the tails of the distribution. Due to the differences in how income and delays are distributed, a similar method has not been considered appropriate in the present analysis.

The estimated values of ρ are shown in Table 49 and range from -0.18 to 0.47 with some of the estimated values being close to 0 and insignificant. The values from the linear dependent variable model are typically lower than those from the ordered logit model. The values are also typically lower if the sample is restricted to delays of up to 15 minutes. When the models are estimated separately for each of the journey purposes, the estimated values of ρ are typically larger for commuters.

To interpret the meaning of ρ estimated from these models, it might be worth providing an interpretation of all the possible values as this parameter indicates the sensitivity of marginal utility to delay. Therefore, a positive ρ suggests that with delay increasing, the marginal utility of delay is becoming less negative whilst a negative ρ suggests the opposite. Therefore if,

1) $\rho > 1$: the marginal utility increases with delay (becomes less negative) at an accelerating rate
2) $0 > \rho < 1$: the marginal utility increases with delay (becomes less negative) at a decreasing rate
3) $\rho = 0$: the marginal utility is constant
4) $0 < \rho > -1$: the marginal utility decreases with delay (becomes more negative) at a decreasing rate
5) $\rho < -1$: the marginal utility decreases with delay (becomes more negative) at an accelerating rate

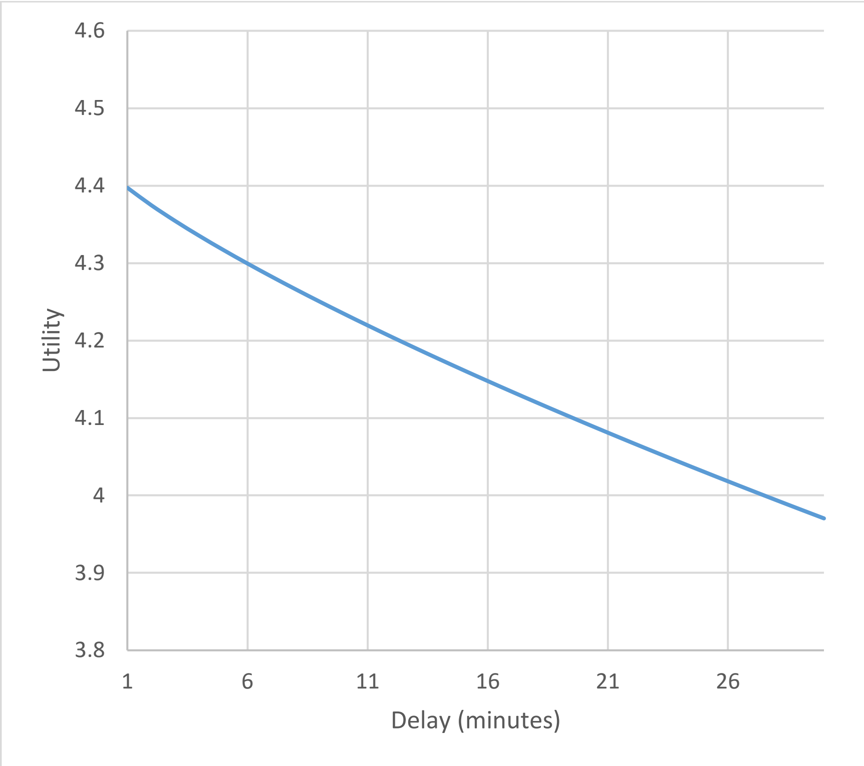
To better understand the difference in the estimated marginal utility of delay for positive and negative ρ, two values were chosen for further investigation. Figure 55 shows the estimated utility and marginal utility using the linear model of overall satisfaction for delays of up to 30 minutes and all the journey purposes combined (i.e. $\rho = 0.184$). Subsequently, Figure 56 depicts the estimated utility and marginal utility using the linear model of overall satisfaction for delays of up to 15 minutes and only leisure travellers (i.e. $\rho = -0.178$).

Table 49 Estimates of ρ

	A (Linear)	B (Linear)	C (Ordered)	D (Ordered)
Overall satisfaction (initial model)				
All	**0.18 [0.1, 0.3]**	0.03 [-0.2, 0.3]	**0.30 [0.3, 0.4]**	**0.15 [0.1, 0.3]**
Business	0.05 [-0.3, 0.4]	-0.05 [-0.9, 0.7]	**0.15 [0.0, 0.3]**	0.04 [-0.3, 0.4]
Commute	**0.20 [0.1, 0.3]**	0.01 [-0.2, 0.3]	**0.36 [0.3, 0.5]**	**0.18 [0.1, 0.3]**
Leisure	0.01 [-0.3, 0.2]	-0.18 [-0.8, 0.3]	0.08 [0.0, 0.2]	-0.11 [-0.4, 0.1]
Overall satisfaction (extended)				
All	0.13 [0.0, 0.3]	-0.11 [-0.8, 0.3]	**0.26 [0.2, 0.3]**	0.07 [-0.1, 0.2]
Business	-0.05 [-0.6, 0.4]	-0.17 [-1.5, 0.8]	0.07 [-0.1, 0.2]	-0.08 [-0.4, 0.3]
Commute	0.11 [-0.1, 0.3]	-0.11 [-0.5, 0.2]	**0.29 [0.2, 0.4]**	0.05 [-0.1, 0.2]
Leisure	0.07 [-0.3, 0.4]	-0.39 [-1.5, 0.7]	**0.18 [0.1, 0.3]**	0.01 [-0.2, 0.3]
Punctuality satisfaction				
All	**0.23 [0.2, 0.3]**	0.10 [0.0, 0.2]	**0.40 [0.4, 0.4]**	**0.26 [0.2, 0.3]**
Business	0.10 [-0.1, 0.3]	-0.14 [-0.5, 0.2]	**0.29 [0.2, 0.4]**	0.03 [-0.2, 0.2]
Commute	**0.32 [0.2, 0.4]**	**0.16 [0.0, 0.3]**	**0.47 [0.4, 0.5]**	**0.31 [0.2, 0.4]**
Leisure	0.07 [0.0, 0.2]	0.00 [-0.2, 0.2]	**0.30 [0.3, 0.4]**	**0.20 [0.1, 0.4]**
Delay lengths	1-29	1-14	1-29	1-14

Legend: *95% confidence intervals in brackets*; estimates significant at 95% shown in bold (models A and B are linear, C and D - ordered logit; the models were estimated using overall or punctuality satisfaction as a proxy for utility; in the extended model, satisfaction with train and station are added as controls)

Interestingly, large differences in the estimated values were observed based on restricting the delay lengths to 15 or 30 minutes what may be suggestive of existence of a point where a maximum dissatisfaction is reached, such that satisfaction data is not able to capture the marginal disutility related to further increases in delay after that point. In such cases, the ρ values changed signs in multiple instances with marginal utility suggested to be increasing (becoming less negative) for models with delay lengths restricted to 30 minutes and decreasing (becoming more negative) for models with delay lengths restricted to 15 minutes. Combining this estimation with previous research on delay perception and dissatisfaction, it is not expected for the marginal utility of delay to increase (become less negative) with delays as this would mean that the very first minute of delay is marginally the worst. Contrarily, it is expected that smaller delays have a lower probability of being perceived and having negative impact on travellers.



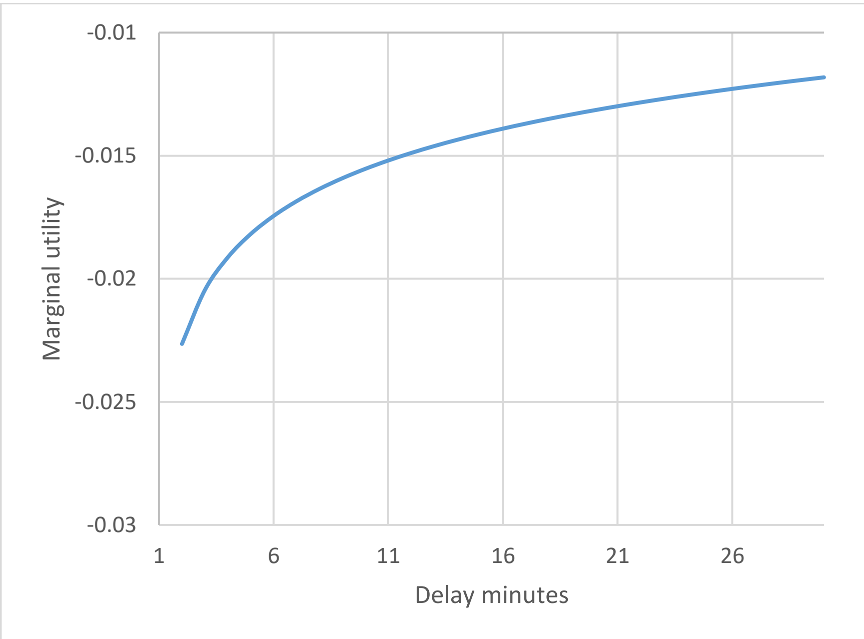
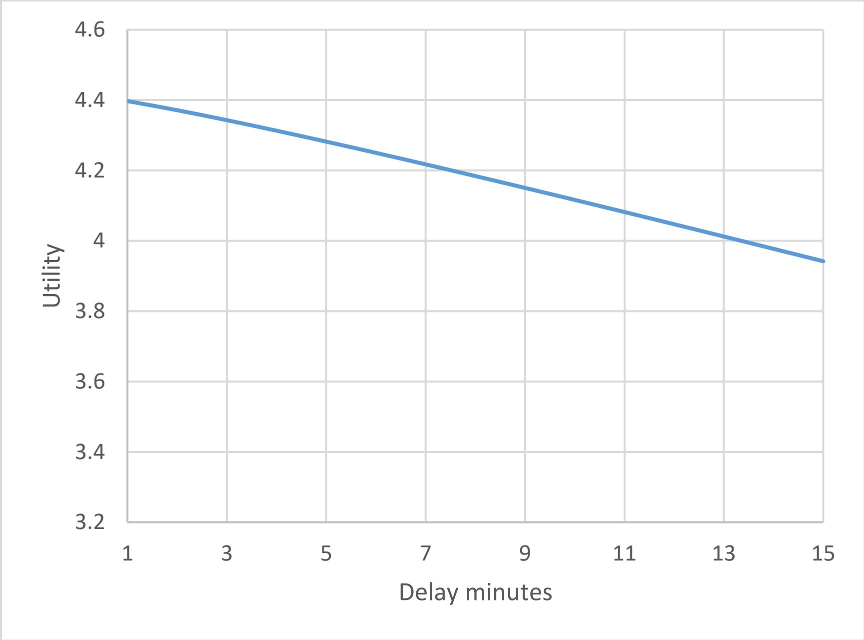


Figure 55 a) Estimated utility for model 1A b) Estimated marginal utility for model 1A



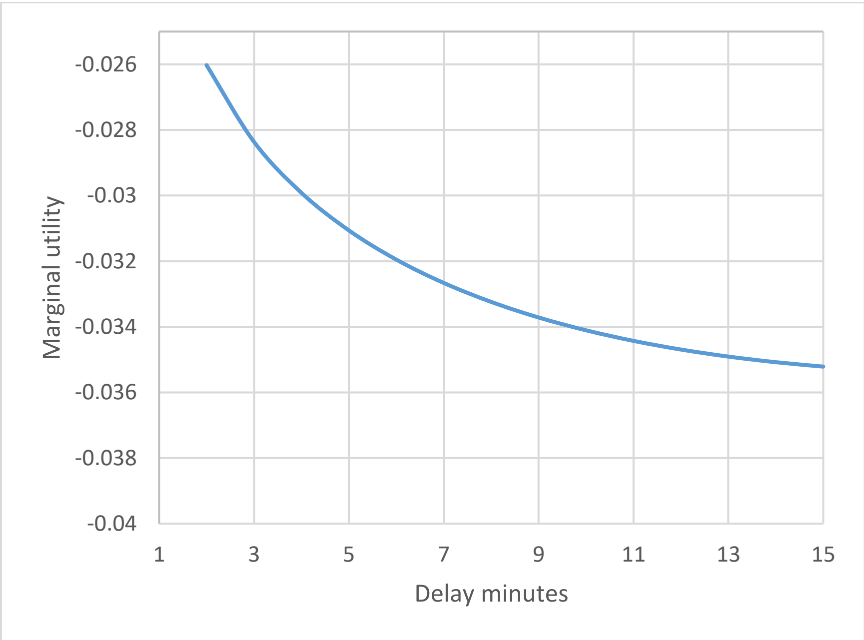


Figure 56 a) Estimated utility for model 1lB b) Estimated marginal utility for model 1lB

Investigating the cubic relationship between delay length and satisfaction

The estimation of the ρ parameter led to both positive and negative values, providing inconclusive evidence as to whether the marginal utility of delay becomes more positive or negative with increasing delays. This, in turn, could suggest that the marginal utility of delay is a non-monotonous function. Hence, estimation of a cubic function (following Gao et al., 2018) was thought to be a particularly useful approach. This was done for:

1) the dependent variable being overall satisfaction (both without and with controls) or punctuality satisfaction (without the additional controls) and
2) restriction of delay lengths to up to 15 or 30 minutes.

The results of the estimated models are shown in Table 50. The predicted delay coefficients are negative for linear and quadratic terms, but positive for the cubic term (though in some cases, some of the coefficients were insignificant or of a different sign), indicating that the marginal impact of delays first becomes more negative and later becomes less negative. Such a function would also lead to marginal utility finally becoming positive, though, this may be related to reaching a point where maximum dissatisfaction is reached (i.e. data imperfections) rather than long delays having a positive marginal impact on utility.

Focusing on the models for delays up to 15 minutes, estimated utilities, marginal utilities and change in marginal utility for models 2a (based on punctuality satisfaction) and 3a (based on the extended model of overall satisfaction) are plotted below.

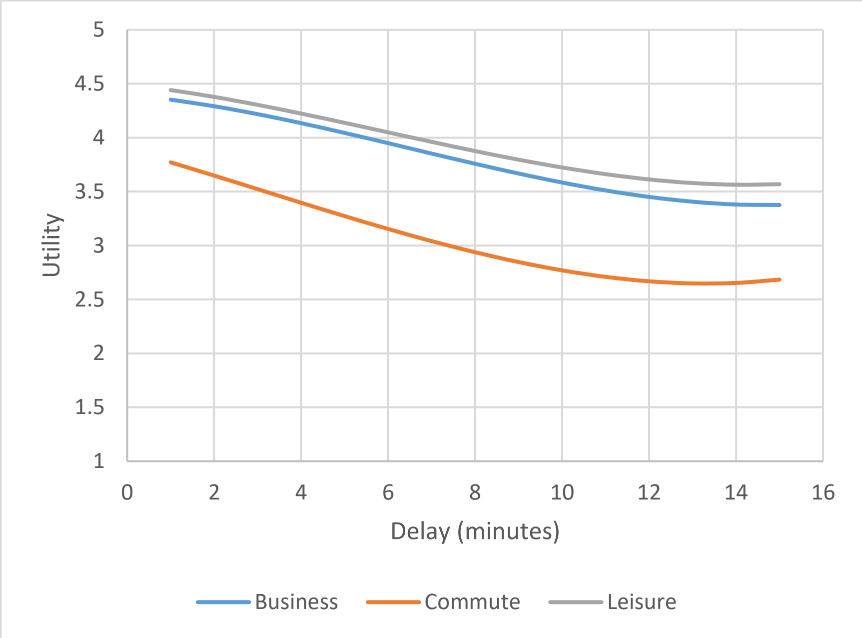
The estimated models indicate that the marginal utility first decreases (i.e. becoming more negative) and then increases (i.e. becoming less negative). This inflection point is estimated to be at around 4 to 10 minutes (depending on journey purpose and the dependent variable used) with both models suggesting on the inflection point being at a lower delay length for commuters.

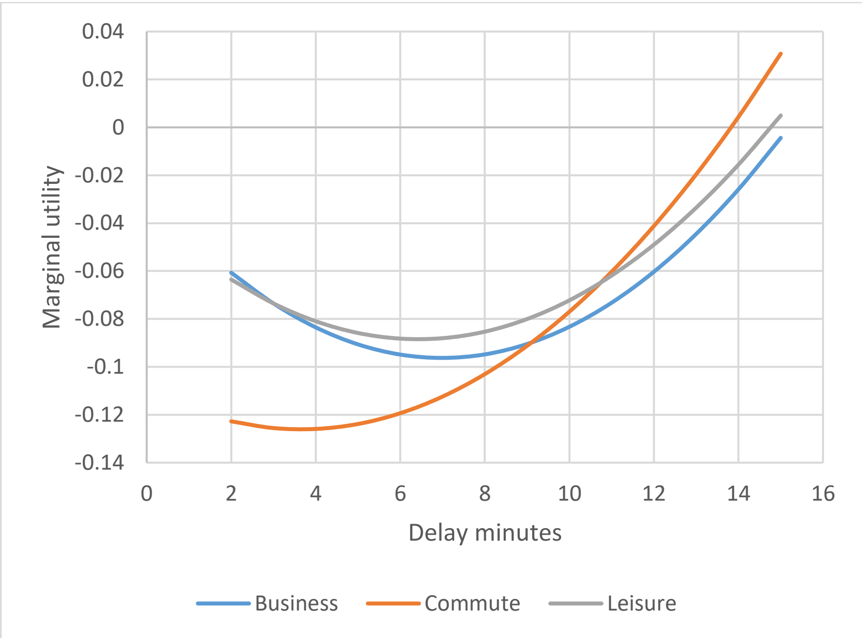
Table 50 Results from a cubic model

	(1)	(2)	(3)	(1a)	(2a)	(3a)
	Overall	Punct	Overall	Overall	Punct	Overall
Constant	4.278***	4.437***	2.942***	4.266***	4.397***	2.944***
	(241.24)	(195.32)	(144.75)	(163.73)	(131.58)	(115.59)
Journey purpose						
Commute	-0.235***	-0.508***	-0.162***	-0.253***	-0.509***	-0.175***
	(-11.40)	(-19.27)	(-7.03)	(-8.38)	(-13.13)	(-6.04)
Leisure	0.146***	0.0798**	0.00996	0.147***	0.0951*	0.0209
	(7.13)	(3.03)	(0.41)	(4.87)	(2.45)	(0.70)
L_A						
Business	-0.0316***	-0.0734***	-0.0197**	-0.0196	-0.0362	-0.00634
	(-3.88)	(-7.03)	(-3.07)	(-1.04)	(-1.49)	(-0.43)
Commute	-0.0711***	-0.153***	-0.0413***	-0.0427***	-0.114***	-0.0264**
	(-13.92)	(-23.46)	(-10.28)	(-3.72)	(-7.77)	(-2.93)
Leisure	-0.0216***	-0.0696***	-0.0182***	-0.0103	-0.0439**	-0.0163

	(-4.35)	(-10.90)	(-4.67)	(-0.90)	(-2.98)	(-1.81)
L_A^2						
Business	-0.000741	-0.00167	-0.000917	-0.00344	-0.00928*	-0.00368
	(-0.87)	(-1.53)	(-1.37)	(-0.99)	(-2.09)	(-1.36)
Commute	0.000503	0.00447***	-0.000763	-0.00529*	-0.00384	-0.00370*
	(0.91)	(6.33)	(-1.75)	(-2.47)	(-1.40)	(-2.21)
Leisure	-0.00134*	-0.00155*	-0.000992*	-0.00378	-0.00753**	-0.00140
	(-2.55)	(-2.29)	(-2.39)	(-1.78)	(-2.75)	(-0.84)
L_A^3						
Business	0.0000294	0.0000838**	0.0000275	0.000184	0.000477*	0.000175
	(1.27)	(2.84)	(1.52)	(1.06)	(2.15)	(1.29)
Commute	0.0000205	-0.0000443*	0.0000422***	0.000321**	0.000406**	0.000188*
	(1.34)	(-2.27)	(3.50)	(2.96)	(2.92)	(2.20)
Leisure	0.0000435**	0.0000730***	0.0000349**	0.000178	0.000424**	0.0000564
	(3.03)	(3.96)	(3.09)	(1.65)	(3.06)	(0.67)
Station_Sat						
Business			0.485***			0.481***
			(31.70)			(30.78)
Commute			0.468***			0.465***
			(54.56)			(53.67)
Leisure			0.472***			0.465***
			(46.52)			(45.03)
Train_Sat						
Business			1.080***			1.064***
			(71.67)			(68.45)
Commute			1.170***			1.166***
			(150.85)			(148.67)
Leisure			1.201***			1.190***
			(118.09)			(113.64)
Delay lengths	1-30	1-30	1-30	1-15	1-15	1-15
N	112992	110369	112992	106196	103710	106196
R^2	0.111	0.173	0.451	0.0924	0.146	0.442

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
L_A refers to delay at arrival; Station_Sat, Train_Sat refer to satisfaction with station and train;
Overall refers to overall journey satisfaction; Punct refers to satisfaction with punctuality.


5
4.5
4
3.5
3
2.5
2
1.5
1
0
2
4
6
8
10
12
14
16
Utility
Delay (minutes)
Business
Commute
Leisure


0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.14
0
2
4
6
8
10
12
14
16
Marginal utility
Delay minutes
Business
Commute
Leisure

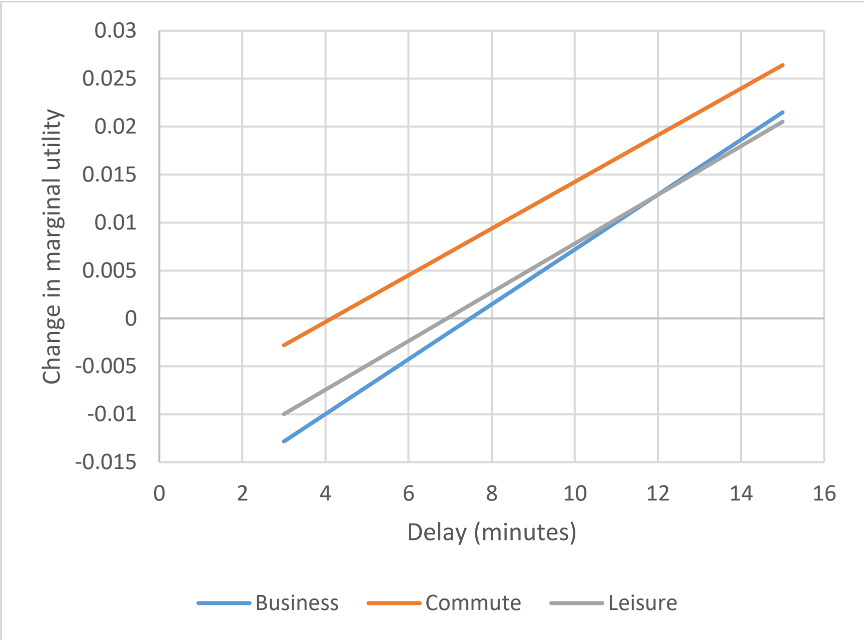
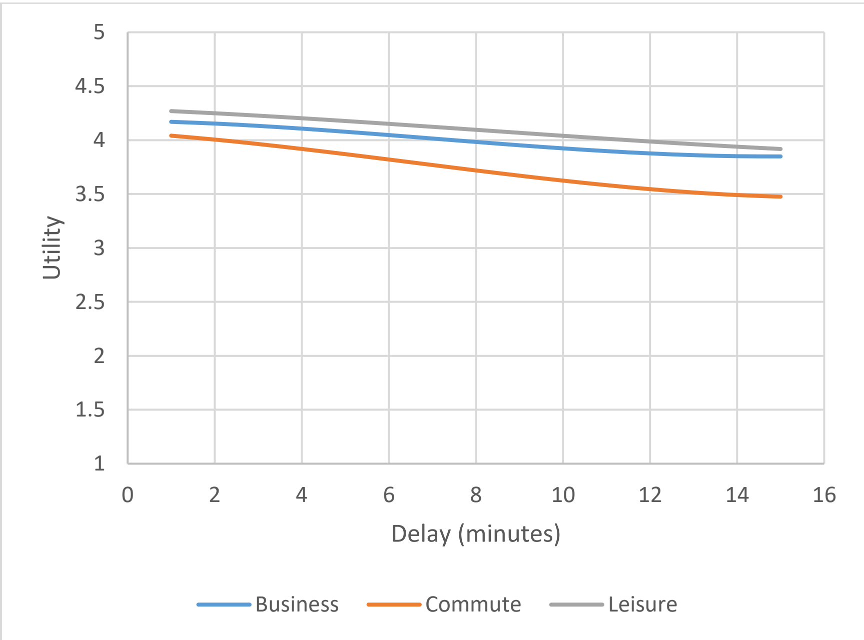
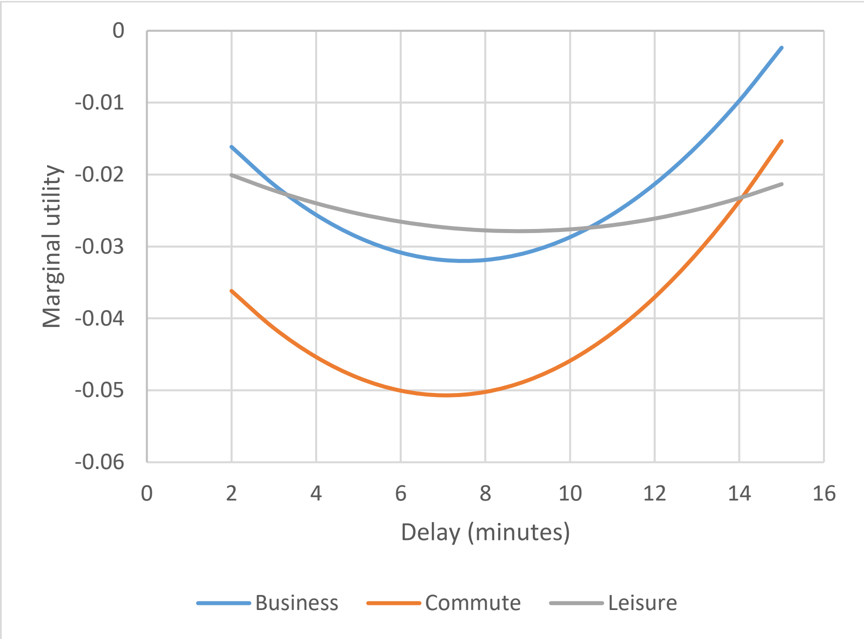


Figure 57 Utility (top), marginal utility (middle) and change in marginal utility (bottom) for model 2a for punctuality satisfaction





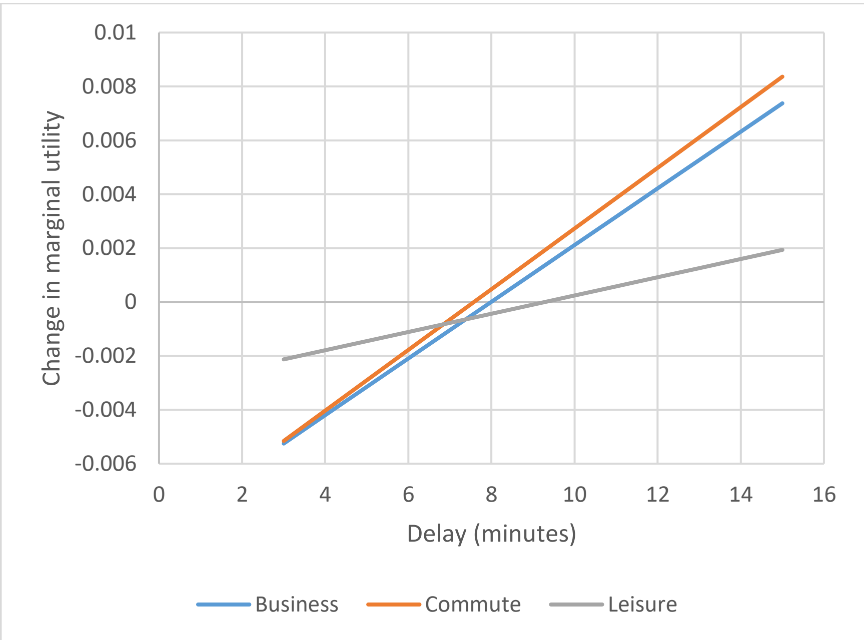


Figure 58 Utility (top), marginal utility (middle) and change in marginal utility (bottom) for model 3a for overall satisfaction

Piecewise regression

Previously estimated models suggested that the marginal utility may be non-monotonous, hence a simpler relationship was estimated based on piecewise regression. This type of regression fits two linear relationships, before and after an estimated breakpoint. Estimating such a breakpoint is in line with suggestions that there exists a delay length where satisfaction data is no longer able to capture the marginal changes in experienced utility, noting that this may be an artefact of the chosen data type. Such a regression was estimated using punctuality satisfaction as the dependent variable with the results reported in Table 51 and the estimated relationship plotted in Figure 59.

Table 51 Results from a piecewise regression

	Business	Commute	Leisure
Constant (β_0)	4.437***	3.887***	4.519***
	(330.15)	(381.57)	(574.58)
β_1	-0.0795***	-0.122***	-0.0768***
	(-36.21)	(-42.75)	(-56.28)
C	19.27***	8.720***	19.48***
	(15.18)	(25.45)	(20.73)
β_2	-0.00615	-0.0430***	-0.0142
	(-0.39)	(-16.14)	(-1.45)
N		113178	
Log-likelihood		-176658.1	
R^2		0.918	

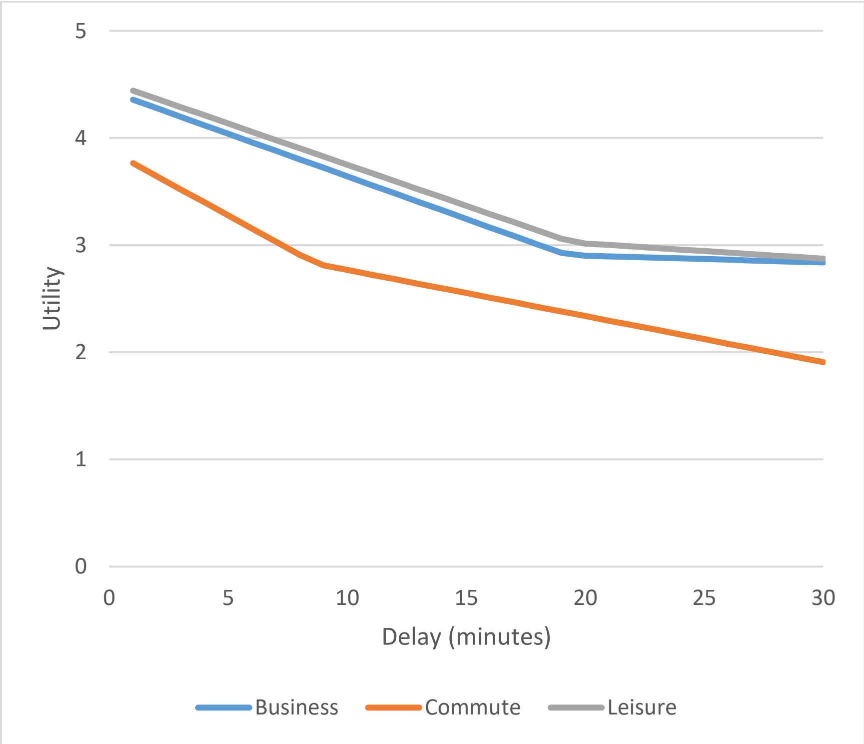


Figure 59 Estimated utility from a piecewise regression

The results obtained from the piecewise regression suggest that the base utility levels are always lower for commuters (as also reported earlier in Chapter 7), in line with the investigation presented earlier in this chapter. Nevertheless, the reduction in utility levels is suggested to be more pronounced for commuters with the break point estimated to be at around 8-9 minutes. Subsequently, the marginal impact of delay is suggested to decrease. For leisure and business travellers, the estimated breakpoint is at around 19-20 minutes. Subsequently, the marginal utility of delay after the breakpoint is not significantly different from 0. It must be noted that the estimated breakpoints are just slightly above the predicted delay dissatisfaction thresholds obtained in Chapter 7. Considering that the marginal utility of delay of business and leisure travellers after the breakpoint is insignificantly different from 0, this raises a question whether there is a point where maximum dissatisfaction is reached. With maximum average satisfaction never reaching 5.0, the small changes in satisfaction after the breakpoint may not be indicative of such delays having no marginal impact. Rather, this may highlight the unsuitability of satisfaction scales to the investigation of the marginal impacts of the longer delays. This may be further complicated by a more limited number of responses related to such delays due to the delay distribution.

7.6.5. Summary

This section aimed to explore the non-linearities in the delay impacts on passengers. Different functional forms were suggested based on previous research and the reported relationship characteristic for the NRPS data. The following steps were taken:

1) Estimation of ρ parameter, allowing to study the curvature of the marginal utility of lateness function based on Layard et al. (2008) work on the marginal utility of income;

2) Estimation of a cubic function describing the relationship between delay length and satisfaction based on Gao et al. (2018);

3) Estimation of a piecewise regression and a break point C determining the length of delay where marginal utility of lateness changes.

The estimation of ρ revealed a range of estimated values, both positive and negative. The values of opposite signs were obtained for the different dependent variables used in the modelling, ordered and linear models as well as for the delays restricted to 15 and 30 minutes. This suggests that a different functional form may be more appropriate for this relationship. Hence, a cubic function was introduced with the predicted delay coefficients being negative for linear and quadratic terms, but positive for the cubic term (though in some cases, some of the coefficients were insignificant or of different signs). This would indicate that the marginal impact of delays first becomes more negative and later becomes less negative. Such a function would also lead to marginal utility finally becoming positive, though this may be due to reaching a point where maximum dissatisfaction is reached. Subsequently, a simpler relationship was studied, assuming that the relationship between delay length and satisfaction is linear up to a certain threshold where it starts following a different linear relationship. The estimated breakpoints were around 9 minutes for commuters and just below 20 minutes for other travellers. It was also suggested that the marginal utility of delay is larger in absolute terms for commuters and subsequently decreases after the breakpoint. However, in the case of other travellers, the marginal utility of delay is not significantly different from 0 after the breakpoint. This was suggested to highlight a couple of limitations of the satisfaction data where:

1) the estimated breakpoints may, in fact, be maximum dissatisfaction thresholds where average reported satisfaction does not reduce anymore - which is not equivalent to such delays having no negative marginal valuation and
2) longer delays being relatively rare – which leads to lower significance of the estimated results for the underrepresented delay lengths

Nevertheless, the work conducted as part of this section introduced different ways of investigating the non-linearities in the impacts of delays on passengers. Some indications of non-constant marginal (dis)utility of lateness were presented, possibly suggesting that the marginal utility of delay is non-monotonous. It may be more negative (and/or becoming more negative) for smaller delays (as the probability of perceiving them increases) and less negative for larger delays. However, the choice of methodology, variables and/or restricting the lengths of delay studied seemed to have a large impact on the estimated results.

7.7. Conclusions

This chapter first introduced the concept of journey (and delay) satisfaction, analysing how travellers' satisfaction changes with increasing levels of delays using various representations of the satisfaction variable (i.e. binary and ordinal) at both individual passenger (i.e. the 'passenger' model) and aggregated (i.e. the 'OD' model) level. This allowed increasing understanding of how incidental and average lateness impact upon passenger satisfaction levels. The estimated models suggested that:

1) Commuters are always less satisfied with a given level of delays and their satisfaction decreases more rapidly.

2) Commuters are already very likely to be dissatisfied even if they face delays of only around 5-10 minutes.
3) For other travellers, these thresholds are typically larger, at around 10-20 minutes.
4) The ‘passenger’ model indicates that the delay incidents of such lengths are detrimental to passenger satisfaction, also highlighting that the impact of delay of a given length is typically lower:
    - if journey quality is better (i.e. passengers are seated);
    - for longer journeys (i.e. suggesting that it is not only the absolute delay length but the delay relative to scheduled journey time that impact passenger satisfaction). This is in line with suggestions by Wardman and Batley (2022), who argued that proportional elasticities (i.e. based on the relative proportion of AML to GJT) better explain changes in demand than the actual delay lengths;
    - with a shorter delay at departure (i.e. travellers are more likely to notice the delays if their train departs late and typically also prefer to be delayed once being on-board as the additional journey time can be used more productively and delay at departure can be related to increased levels of uncertainty).
5) The predicted delay length dissatisfaction thresholds are indicated to be lower than previously suggested in the literature (i.e. Monsuur et al., 2021).
6) The ‘OD’ model suggested that commuter-focused OD pairs are likely to suffer from lower levels of satisfaction as:
    - under perfect performance, 80% of commuters and 90% of other travellers are suggested to be satisfied with performance and
    - the reduction in the proportion of satisfied passengers is more profound for commuters - (on average) 6 pp for a minute of APL for commuters and 2 pp for other travellers.

This analysis then suggested that the estimated models could be used in setting performance regimes and/or targets or help design passenger compensation mechanisms where:

1) the ‘passenger’ model can be used to determine the preferable distribution of delays (i.e. minimising the incidence of delay episodes that are detrimental to passenger satisfaction),
2) the ‘OD’ model at the same time can be used to determine the preferable average performance targets and

3) it should be noted that a combination of both approaches (i.e. lowering the incidence of longer delays and average delays simultaneously) is likely to achieve the best results in terms of passenger satisfaction. This means that the focus should not only be on minimising the average lateness, but also reducing the incidence of the longer delays.

However, it has been noted that due to the non-quantitative nature of the satisfaction scores, the classification of satisfaction versus dissatisfaction is subject to interpretation. This also means that it is not immediately clear what the differences between the satisfaction categories mean and how various levels of satisfaction translate to well-being. This is one of the key limitations of using satisfaction data in setting policy targets. As suggested throughout this chapter, depending on the assumption related to how satisfaction versus dissatisfaction is classified, the targets may differ. Nevertheless, V2 of the binary representation provides the way of classifying satisfaction versus dissatisfaction that best aligns with the conceptual framework of studying the threshold of positive/non-positive satisfaction. Ultimately, how different satisfaction levels may affect wellbeing and rail demand is more difficult to discern. Regardless of the demand impacts, however, studying satisfaction may allow to understand the negative impacts of delays that occur even in the case where the worsening performance does not translate to demand impacts.

Finally, this chapter reconciled the concepts of delay perception and satisfaction to test the hypothesis that there is a gap between the length of delay that is perceivable and such that impacts satisfaction. It was suggested that:

1) there is a gap between delay perception and dissatisfaction as the probability of perceiving a delay is larger than the probability of being dissatisfied given the same length of delay incident,
2) the estimated gap is suggested to be between 0 minutes for standing commuters to up to 9 minutes for seated leisure travellers and
3) it has been generally suggested that the gap between delay perception and dissatisfaction increases for seated passengers, with longer journeys and delay at departure.

This chapter focused on establishing a link between delay length and passenger satisfaction with the use of logistic regression. The concepts of delay perception and satisfaction were introduced, suggesting that smaller delays are often unperceived and have a small impact on passenger satisfaction. This, in turn, raised a question about the curvature of the shape of the relationship between delay and travellers' utility. Such concerns were generally raised in the literature, e.g. by Wardman and Batley (2022) arguing that proportional elasticities (i.e. based on the relative proportion of AML to GJT) better explain changes in demand than the actual delay lengths. Hence, an attempt was made to explore the potential non-linearities in the delay impacts using the satisfaction data from NRPS (as a proxy of individual utility) by applying approaches introduced by previous literature. Different functional forms were suggested based on previous research and the reported relationship characteristic for the NRPS data. Whilst the results were inconclusive, the work conducted as part of this chapter introduced different ways of investigating the non-linearities in the impacts of delays on passengers. Some indications of non-constant marginal (dis)utility of lateness were suggested by the estimated models, possibly indicating on the marginal utility of delay being non-monotonous. It may be more negative (and/or becoming more negative) for smaller delays (as the probability of perceiving it increases) and less negative for larger delays. However, the choice of methodology, variables and/or restricting the lengths of delay studied seemed to have a large impact on the estimated results. Moreover, the analysis suggested that one of the limitations of the satisfaction data in this context may be related to the existence of a point where the 'maximum dissatisfaction' is reached such that it is difficult to establish the relationship for longer delays.

There are multiple ways in which the analysis conducted here could be improved. First of all, additional segmentation could be introduced based on journey lengths, geographies and/or sociodemographic characteristics. Moreover, since the focus of this analysis was on the absolute lengths of delays, a relative approach could be explored instead. In this case, the delay could be represented as the proportion of scheduled journey time or GJT. This would be more in line with the suggestions by Wardman and Batley (2022) where the demand elasticities based on relative proportions of AML to GJT were indicated to better explain changes in demand. However, it has to be noted that there are important differences between AML and incidental delays that are related to the type of data used (as discussed in Chapter 5).

Future studies could also explore the non-linearity of delay impacts using the methodologies outlined in this chapter for different satisfaction surveys. It might be particularly useful to see if similar results are obtained for surveys from different countries, different modes or other satisfaction scales.

Having compared the concepts of delay perception and dis(satisfaction), the purpose of the last empirical chapter of this thesis is to use satisfaction data to estimate lateness multipliers, defining a trade-off between a minute of delay to an equivalent length of scheduled journey time.

Chapter 8

Lateness valuation using satisfaction data

8.1. Introduction

Transport researchers are interested in the impacts that different journey aspects, including scheduled journey times, fares, delays and comfort have on rail passengers. These are often evaluated using demand data (e.g. Wheat and Wardman, 2017), stated preference (e.g. Ibáñez, 2012) or satisfaction surveys (e.g. Monsuur et al., 2021). Regardless of the source of the data, there generally is a consensus in the literature that delays negatively impact transport users, affecting both their satisfaction and travel choices. However, it has been suggested that the observed changes in demand in response to worsening performance (estimated in the market-level econometric analyses) are relatively limited as compared to the lateness valuation derived from individual-level discrete choice studies (Batley et al., 2011).

Several studies attempted using Stated Preference (SP) data in delay valuation, where the so-called lateness multipliers (also referred to as reliability multipliers by some authors) define the conversion rate between 1 minute of lateness to the equivalent of journey time and in this sense are defined as the trade-off between lateness and scheduled journey time (e.g. Börjesson and Eliasson, 2011; Batley and Ibáñez, 2012). In the British context, most studies supported the lateness multipliers of around 3 - i.e. 1 minute of lateness being valued as the equivalent of 3 minutes of scheduled journey time (for review see Wardman and Batley, 2014). This chapter draws on earlier work using SP surveys to estimate lateness multipliers (e.g. Bates et al., 2001; Preston et al., 2009; Börjesson and Eliasson, 2011; Wardman and Batley, 2022 and particularly Batley and Ibáñez, 2012) whilst estimating the lateness multipliers using journey satisfaction data. At the same time, the methodology used in this study is similar to the large body of literature using data from surveys on life satisfaction (e.g. Layard et al., 2008; Dickerson et al., 2014). The major difference is the use of a survey on journey, not life, satisfaction and its cross-sectional nature. Building on the work by Monsuur et al. (2021) and the previous chapter, the National Rail Passenger Survey is used to estimate an ordered logit model of passenger satisfaction and estimate the utilities of both scheduled journey time and delay (at departure and arrival) for a pseudo-panel of frequent rail travellers. The estimated coefficients are subsequently used in the estimation of lateness multipliers. Hence, the overall aim of this work is to use a dataset that is novel in the context of lateness valuation (i.e. responses from a survey on journey satisfaction) and apply it to the established methodologies to:

1) explore the potential of journey satisfaction data in economic valuation in transport-related contexts and
2) compare the lateness multipliers estimated from satisfaction data to the values obtained from the traditional methods (i.e. SP surveys).

This chapter is structured as follows:

- section 8.2 presents a literature review, positioning the lateness multipliers within the British rail forecasting framework as well as a description of the data sources typically used in their estimation,
- section 8.3 describes the data and the modelling approach undertaken,
- section 8.4 presents the estimated models and lateness multipliers,
- section 8.5 provides a comparison of the estimated values with previous literature,
- section 8.6 summarises this work and discusses the potential for using satisfaction surveys in the future research concerning economic valuation in transport.

8.2. Literature review

8.2.1. Lateness valuation framework

Ticket sales data is often used to estimate the effect that Generalised Journey Time (GJT) components have on rail demand (for a review see Wheat and Wardman, 2017). Following Wheat and Wardman (2017), the rail demand function in Great Britain (GB) is specified as:

$$V = \mu GJT^{\lambda} F^{\gamma} GVA^{\delta} \tag{29}$$

where GJT is generalized journey time, F is fare, GVA is income, λ, γ, δ are the respective elasticities and μ represents all the other factors impacting the demand. Generalised journey time in this formulation is a composite index specified as:

$$GJT = T + \alpha H + \beta I \tag{30}$$

where T is the station-to-station journey time, H is a service headway and I is the number of interchanges with α and β being the respective penalty multipliers converting both the number of interchanges and service headway into equivalent journey time.

Extending the demand specification, Batley et al. (2011) used the following relationship between demand and average lateness at the destination, previously prescribed by passenger Demand Forecasting Handbook (PDFH) in Great Britain (ATOC, 2004):

$$Y = \left[1 + \frac{w(\bar{L}^{+}_{new} - \bar{L}^{+}_{base})}{GJT_{base}}\right]^{\lambda} \tag{31}$$

where Y is the proportionate change in rail demand, $\bar{L}^{+}_{new}$ and $\bar{L}^{+}_{base}$ represent average lateness at the destination in the new and base scenarios, GJT_{base} is the generalized journey time in the base scenario, λ is the elasticity of rail demand to generalized journey time and w is the reliability (lateness) multiplier.

As noted by Wheat and Wardman (2017), PDFH is a set of guidelines and forecasting parameters that combine years of research into rail demand in Great Britain, providing a comprehensive and consistent framework for economic appraisal of railway schemes. Of note, as discussed by Wardman and Batley (2022), this represents the so-called 'indirect' approach to forecasting the impact of changes in railway performance. Since 2018 (PDFH v6), a recommendation was made to move to a 'direct' approach where a change in demand Y is estimated directly based on a change in average lateness and the late time elasticities (usually obtained from rail demand models, for review see Wardman and Batley, 2014).

The aforementioned lateness multiplier w defines the conversion rate of 1 minute of lateness to the equivalent of journey time and in this sense is defined as the trade-off between lateness and scheduled journey time. It is typically estimated as the ratio of the utility of lateness to the utility of scheduled journey time. Wardman and Batley (2014) provide a review of estimates of lateness multipliers since 1984 with most of the initial values being around 3. Similar studies conducted throughout the years generally supported that figure but suggested values of up to 6.5 for airport journeys with Batley and Ibáñez (2012) estimating lateness multipliers for different demand segments based on journey purposes and lengths as shown in Table 52. As reported in Wardman and Batley (2014), in most cases, the estimated lateness multipliers range between 2-5 for business travellers and commuters and 2-7 for leisure travellers, though values larger than 10 have also been reported throughout the literature (e.g. Wardman, 2001; Börjesson and Eliasson, 2011).

Table 52 Lateness multipliers (Batley and Ibáñez, 2012)

Journey Purpose	Short	Long
Business	2.68	1.78
Commute	3.12	2.00
Other	5.19	1.77

8.2.2. Data sources used in the estimation of lateness multipliers

Stated Preference (SP) surveys are most often used in studies where lateness multipliers are estimated (e.g. Bates et al., 2001; Preston et al., 2009; Börjesson and Eliasson, 2011; Li et al., 2016). In such cases, passengers are presented with alternative hypothetical travel options and make a choice regarding their preferred scenario. The differences in the options presented to the respondent are the ticket prices, scheduled journey times and performance (presented as average delay or distribution of delays). An example of such an approach is Batley and Ibáñez (2012) where one of the pairs of journey options shown to respondents was:

1) Option A where a 27-minute journey cost £2.40 with average lateness of 1 minute at departure and 4.4 minutes at arrival.
2) Option B where a 23-minute journey cost £3.60 with average lateness of 4.4 minutes at departure and 8.8 minutes at arrival.

While the SP data can be subjected to biases, such as systematic bias (divergence between hypothetical and actual choices), justification bias (rationalizing actual choices) or strategic bias (influencing policy) (for review see Wardman, 1988), it has become a standard approach. Indeed, it is often the only possible source of such data (Bates et al., 2001) as SP studies allow the analyst to design scenarios that may not be observable in the real world as well as explicitly control for the choice attributes (Tsoleridis et al., 2022). An alternative to stated preference data is revealed preference (RP) data where passengers' actual travel choices are investigated. While economists typically prefer data on actual choices, the RP data has its own limitations. It is more difficult to obtain, may be prone to reporting errors (especially in the case of traditional travel diaries) and is based on the assumptions of perfect information about the possible travel alternatives whereas, in fact, it is difficult to identify the choice sets and trade-offs faced by the participants (Wardman, 1998; Bates et al., 2001; Hess et al., 2007; Preston et al., 2009; Tsoleridis et al., 2022).

An alternative to SP and RP surveys can be sought in satisfaction surveys where passengers score their satisfaction with an actual travel experience *ex-post*. There is an abundance of literature looking at the impact of different journey aspects on passenger satisfaction (for reviews see De Vos et al., 2013; De Oña and De Oña, 2015; Gao et al., 2018; Ye et al., 2022). Unlike SP or RP studies, passengers are not faced with multiple alternatives but score their satisfaction with a particular journey (though some studies analysed surveys referring to general satisfaction with public transport, e.g. Cats et al., 2015). Most studies cite travel time, monetary cost, performance, journey comfort and provision of information as key determinants of passenger satisfaction (e.g. Brons and Rietveld, 2009; Carrel et al., 2016; Börjesson and Rubensson, 2019; Lunke, 2020; Monsuur et al., 2021). In the British rail context, Monsuur et al. (2021) used the National Rail Passenger Survey to estimate the impact of delays on passenger satisfaction, suggesting that passengers are very unlikely to remain satisfied with journeys delayed by over 30 minutes, also highlighting the importance of journey quality on travel satisfaction.

When considering satisfaction survey data, one must also remember about the limitations. As with SP surveys, respondents may be biased, aiming to influence the results (strategic bias). Furthermore, travellers may have imperfect knowledge or may not recall the correct answers to responses in the surveys (Choi & Pak, 2005). With NRPS, for example, it is highlighted that the timing of handing out the questionnaire (i.e. prior to boarding a train) may also impact when and how the travellers respond, hence affecting the results. It is likely that satisfaction surveys also are an imperfect source of data. That said, given that the sources of these imperfections are different to those encountered with SP or RP surveys, satisfaction surveys are a useful addition, offering an alternative and/or complementing those more traditional data sources.

Whilst data on scheduled journey time and lateness may be available to supplement the reported satisfaction, it refers to incidental (i.e. for a specific journey), not mean or standard deviation of performance, as is typically the case with SP surveys. Satisfaction data, typically from longitudinal household panels, have been used in economic valuation in labour (e.g. Layard et al., 2008), health (e.g. Ferrer-i-Carbonell and van Praag, 2002) and environmental economics (e.g. Frey et al., 2009). However, similar approaches have not been as widely used in transport economics, possibly resulting from a lack of transport surveys with such detailed information or from household surveys lacking enough transport-related information. The most important exception is a study by Dickerson et al. (2014) looking at the relationship between life satisfaction and commuting. Hence, in this context, the use of journey satisfaction data in the estimation of lateness multipliers represents a relatively novel approach. The following section provides a more detailed description of the proposed approach.

8.3. Methodology

8.3.1. Pseudo-panel of frequent travellers from NRPS

NRPS dataset described in Chapter 5 was used in the analysis conducted as part of this chapter. Following the initial analysis presented in Chapter 5, modelling of delay perception in Chapter 6 and passenger satisfaction in Chapter 7, the dataset was further restricted.

With the NRPS dataset being cross-sectional in nature, an attempt was made to create a subset of the original dataset capturing frequent rail travellers. Pseudo-panel approaches have been widely used in the literature in the absence of true panel datasets (e.g. Dargay, 2002; Rich et al., 2023). This pseudo-panel of frequent travellers was then used to investigate the impact of both scheduled journey time and delays on passenger satisfaction.

It is expected that while delays affect overall journey satisfaction of both frequent and infrequent travellers (i.e. as investigated in Monsuur et al., 2021), the scheduled journey time itself should not generally directly impact satisfaction with an individual journey. This is based on an assumption that travellers' decision to travel on a given service characterised by a timetabled scheduled journey time was one that maximised travellers' utility. However, as suggested by Cats et al. (2015), longer journeys may be associated with lower overall satisfaction with public transport for commuters.

This can be illustrated using an example of two journeys:

1) A long-distance 4-hour business journey between London and Edinburgh
2) A short-distance 30-minute commuter journey between London and Stevenage

The two examples presented above represent very different journeys. Considering the cross-sectional nature of the dataset, it is not believed that the differences in the timetabled lengths of journeys have an impact on journey satisfaction of the two different types of travellers. Journey time is a source of disutility for passengers travelling on different origin-destination (OD) pairs, but assuming that both types of travellers are rational and aim to maximise their utility, the choices to travel from London to Edinburgh and from London to Stevenage are ones that maximise their utility. Assuming that both services perform as timetabled, it would be expected that both types of travellers are satisfied with their journeys. Any differences in the satisfaction scoring may be due to the differences in other journey aspects (e.g. comfort) that may be correlated with journey length.

On the other hand, a frequent traveller, i.e. between London and Stevenage may be able to perceive changes in scheduled journey times (timetable). Such changes may, in turn, affect their journey satisfaction. Investigating this relationship is, however, only possible for panel (not cross-sectional) datasets where the same traveller scores their satisfaction with multiple different journeys on the same OD pairs across time. Hence, constructing a pseudo-panel of frequent travellers allows investigating whether and how the changes in scheduled journey times and experienced lateness on the same OD pair affected reports of journey satisfaction.

To align with this framework, a number of modifications was applied to the original NRPS dataset, based on the following journey characteristics:

1) Frequency of travel

Out of the 46% of passengers responding to the question regarding the frequency of travel on a given route, 73% admitted to travelling at least every 2 months (classified as frequent travellers for the modelling purposes). It is assumed that delays may affect satisfaction of both frequent and infrequent travellers. However, it is only the frequent travellers whose satisfaction is assumed to be affected by potential changes in scheduled journey times on a given route.

2) Recorded delay length and delay perception

Responses where a passenger reported late arrival but no delay was matched using the operational data (5.7%) were discarded, as were responses associated with delays of more than 30 minutes – so as to remove outliers and possibly erroneous responses as in some cases large differences between recorded and reported delays were found for these records. However, responses where no delay was reported and recorded were retained as the interest here is in both delayed and on-time journeys (similarly to modelling journey satisfaction in Chapter 7).

3) Number of responses for a given origin-destination (OD) pair

OD pairs with more than 10 and 25 responses were selected. 792 OD pairs were identified with more than 10 responses (over 26,026 records) and 270 pairs with more than 25 responses (over 17,695 records). The response thresholds of 10 and 25 were selected arbitrarily to find a compromise between the number of OD pairs and the number of responses per OD.

In conclusion, between 14,000 and 40,000 responses are used in the estimation of the satisfaction models described in the following section. This depends on the choice of OD pairs as well as control variables (i.e. the more control variables, the fewer responses as some questions in the survey were subject to non-response).

Passengers scored their overall satisfaction with journeys on a 5-point Likert scale, from ‘very satisfied’ to ‘very dissatisfied’ as discussed in Chapter 5. Unlike in most of the analysis presented as part of Chapter 7, here the overall journey satisfaction score (instead of punctuality satisfaction) is used as the dependent variable as the interest lies in understanding the impacts of both delay and scheduled journey time on satisfaction. Punctuality satisfaction used in Chapter 7 is not well-suited to this body of analysis as it is expected to be only affected by lengths of delay (but not scheduled journey time, at least directly), not allowing to estimate a trade-off between delay and scheduled journey time. Similarly, passengers scored their satisfaction with other journey aspects, e.g. train, station, value for money and service frequency, as discussed further in the following sections. These are used to control overall journey satisfaction for satisfaction with other journey aspects.

8.3.2. Deriving the lateness multipliers

As the dependent variable (overall journey satisfaction) can take one of the five outcome categories, which are in sequential order, an ordered logit model is used for estimating the latent continuous variable y*. In this case, the probability of choosing a satisfaction category i is estimated for a given number of k threshold (i.e. 4 thresholds given 5 satisfaction categories), thus:

$$P(Y = i) = P(k_{i-1} < y^* \leq k_i)$$

(32)

where journey satisfaction is modelled as follows:

$$P(Y = i) = P(k_{i-1} < \beta_0 + \beta_1 SJT + \beta_2 L_D + \beta_3 L_A + \sum_{n=1}^{n} \beta_n Sat_n + \sum_{j=1}^{j} \gamma_j OD_j \leq k_i)$$

(33)

where:

SJT: scheduled journey time

L_A: length of delay at arrival (destination)

L_D: length of delay at departure (origin)

OD_j: OD pair j

Sat_n: a dummy for a variable representing passenger’s satisfaction with train and station (models 1-4) and also satisfaction with value for money and frequency (model 4). It takes the value of 1 if a passenger was ‘very satisfied’ or ‘fairly satisfied’ with a given journey aspect or 0 otherwise. It is important to acknowledge the potential endogeneity bias. This is related to needing to control the overall journey satisfaction for those specific journey aspects that are difficult to measure and/or there is no alternative variable that could control for their effect on the overall satisfaction.

In principle, the proposed model is very similar to the punctuality satisfaction models estimated in Chapter 7. However, in models 2-4, OD pair fixed-effects (Baltagi, 2021) are included by introducing a dummy variable representing each of the OD pairs represented in the sample. This allows the treatment of the dataset, which strictly speaking is cross-sectional in nature, as a pseudo-panel of frequent rail travellers, to estimate the impacts of both changes in scheduled journey times and delays on passenger satisfaction.

Ordered logit model is conceptually more suitable for modelling ordinal data than linear regression (Dickerson et al., 2014; for a review see Boes and Winkelmann, 2006), but its major disbenefit is the difficulty in directly interpreting the coefficients. However, as noted by Dickerson et al. (2014), the ratios of the coefficients in the ordered model can be used to evaluate the trade-offs between variables. In this case, lateness multipliers are estimated as a ratio of the utility of departure and arrival delay $\beta_2$and β_3 to the utility of scheduled journey length β_1. The multipliers are calculated separately for the two types of delays, at departure (w_D) and arrival (w_A) following Batley and Ibáñez (2012) for the selected three journey purposes. In line with the literature (i.e. Bates et al., 2001; Preston et al., 2009; Batley and Ibáñez, 2012), the lateness multiplier represents the value of delayed time with respect to the scheduled time:

$$w_D = \frac{\beta_2}{\beta_1} \text{ and } w_A = \frac{\beta_3}{\beta_1} \tag{34}$$

8.3.3. Choice of control variables and demand segmentation

In terms of the choice of control variables, as noted previously, passengers' satisfaction with public transport is not only impacted by performance but also by other journey aspects. Chapter 5 described the different satisfaction variables forming part of the NRPS dataset. However, given the similarity in the questions, relatively high non-response rates for some of them as well as correlations, the choice of the control variables was limited to the two general satisfaction questions - satisfaction with train and station as well as satisfaction related to timetable or performance, i.e. satisfaction with train frequency, punctuality, scheduled journey time and value for money. However, satisfaction with punctuality and scheduled journey time are represented in the model by the directly observed (experienced) values, which is generally a preferred approach as using satisfaction variables as explanatory variables in the case of the overall satisfaction models introduces endogeneity bias.

Initially, an attempt was made to directly compare the results obtained by estimating lateness multipliers using the methodology described in section 8.3.2 and the journey type categorisation used in Batley and Ibáñez (2012). The same 12 OD pairs were chosen and categorised as long and short distance following Batley and Ibáñez (2012) as shown in Table 53. The decision to choose the same OD pairs as well as to follow the original journey type classification was made to allow more direct comparison of the results with the original research what was considered a sensible initial step, given the novelty of the proposed approach. With only 1176 responses across these pairs, the obtained model results were mostly insignificant. This is perhaps due to the fact that while in SP surveys the choice is fully determined by the described attributes and preferences, in the case of satisfaction surveys, the choice of passenger satisfaction scores is much more complex. Focusing on a smaller subset of OD pairs leads to a reduced number of responses and, in turn, with the distribution of delays typically skewed towards smaller values, only 248 out of 1176 responses were characterised by delays of over 5 minutes – 61 for long distance and 187 for short distance OD pairs. These numbers are reduced further following journey purpose categorisation (as shown in Table 55), as well due to non-responses to some of the satisfaction questions described in Chapter 5.

Table 53 Journey type OD categorisation based on Batley and Ibáñez (2012).

Long distance	Short distance
Bristol-London	Brighton-London
Leeds-London Kings Cross	Glasgow-Edinburgh
Swindon-London Paddington	Leeds-Sheffield
Leeds-Birmingham New Street	Peterborough-London Kings Cross
	Portsmouth Harbour-London Waterloo
	Reading-London Paddington
	Stevenage-London Kings Cross
	Woking-London

The initial model employs segmentation by three journey purposes - business, leisure and commute whilst the segmentation in the extended model aligns with that used by Batley and Ibáñez (2012) (i.e. by three journey purpose and two journey type categories) for better comparison of the estimated values. Therefore, all the responses were categorised based on the journey type classification provided in the dataset by Transport Focus and subsequently grouped as short and long, as shown in Table 54 to align with Batley and Ibáñez (2012). This classification is based on the genre definition used for the segmentation of the different services into 7 building blocks. As noted by Transport Focus, this classification aligns with operational data for sub divisions of the TOCs’ networks and is used to benchmark performance against the respective building block genre (Transport Focus, 2020).

Generally, commuter type services have been classified as short whereas the high-speed, interurban and long distance services were classified as long journeys with details on the distribution of responses provided in Table 55 and summary statistics presented in Table 56. It is noted that this provides only one of many possible ways to segment the data. The alternative split could be based on journey length or distance. Finally, it is noted that the proposed segmentation not controls for the journey lengths, but also service types which can be related to the speed/distance ratios as well as prices of the different services forming part of the different NRPS building blocks. It is hypothesised that this allows for additional insights related to how different travellers value different aspects of their journeys.

Table 54 OD pair distribution across journey type categories

NRPS building block	OD pairs	Journey type category	Average SJT
Airport	11	-	24
High-speed	37	Long	82
Interurban	37	Long	55
Long commute	85	Short	45
Long distance	45	Long	91
Rural	11	-	50
Short commute	44	Short	31

Table 55 OD pair distribution across journey purposes and journey types

Distance	**Purpose**			
	Business	Commute	Leisure	Total
Short (%)	6.75	24.40	14.56	45.72
Long (%)	18.33	9.86	26.10	54.28
Total (%)	25.08	34.26	40.66	100.00

Table 56 Descriptive statistics corresponding to model 4a from Table 59

Segment	Overall	Station	Train	Freq	VfM	L_A	L_D	SJT	N
SB	4.14	0.86	0.84	0.87	0.43	2.68	1.39	58.7	947
SC	3.82	0.80	0.72	0.75	0.20	2.16	1.31	35.3	3,504
SL	4.36	0.90	0.88	0.90	0.62	1.95	1.14	51.1	2,068
LB	4.15	0.84	0.84	0.88	0.44	3.68	1.31	103.5	2,556
LC	3.77	0.80	0.71	0.79	0.20	3.92	2.62	37.9	1,418
LL	4.39	0.88	0.90	0.91	0.67	3.10	1.44	99.7	3,653

Overall refers to overall journey satisfaction; Station, Train, Freq and VfM are the proportions of passengers reporting satisfaction with station, train, frequency and value for money; L_A: delay at arrival and L_D: delay at departure; SJT: scheduled journey time; N: number of responses by demand segment; SB/LB – Short/Long Business, SC/LC – Short/Long Commute, SL/LL – Short/Long Leisure;

8.3.4. Summary

Three iterations of models were estimated using Stata 17 (StataCorp, 2021) and are presented along three sets of the estimated lateness multipliers. For each of the iterations, four alternative specifications of the models are presented. The simpler versions of the models were estimated with a more limited segmentation whereas the extended models followed the segmentation similar to that in Batley and Ibáñez (2012). Additionally, as part of a sensitivity analysis, models with the exclusion of departure delay were estimated. In all cases, four sets of models are presented:

1) Ordered logit model without OD fixed effects;
2) Ordered logit model with OD fixed effects for a subset of OD pairs with at least 10 responses;
3) Ordered logit model with OD fixed effects, for a subset of OD pairs with at least 25 responses;
4) Ordered logit model with OD fixed effects, for a subset of OD pairs with at least 25 responses and additional control variables representing satisfaction with value for money VfM_{Sat} and service frequency $Freq_{Sat}$.

8.4. Results

The initial models

The models of passenger satisfaction were estimated using an ordered logit model with estimated coefficients presented in Table 57. Model 1 is based on estimating the ordered logit without OD fixed effects. In this case, the delays at arrival and departure both have a statistically significant negative impact on satisfaction whilst the impact of scheduled journey time is less clear. As discussed previously, it is not expected for passengers travelling on different OD pairs to be less satisfied with the longer journey. Using overall journey satisfaction rather than punctuality satisfaction as the dependent variable means that the satisfaction needs to be controlled for other journey aspects. This is represented by the positive impact that satisfaction with station and train are suggested to have on the overall satisfaction. However, it is worth noting the significant and negative coefficient on scheduled journey time under model 1 for commuters that may be potentially explained by commuters generally showing larger dissatisfaction with longer travel for work as indicated by Cats et al. (2015). Nevertheless, it can be expected that respondents who travel on the same OD pair are sensitive to changes in scheduled journey times and it is further assumed that these impacts are similar for travellers on the same OD. With the introduction of OD fixed effects in models 2-4, the coefficients on scheduled journey time become significant and negative for all journey purposes, in line with expectations.

It is noted that the models where the outcome variable is overall satisfaction with the journey may suffer from endogeneity bias. As satisfaction with a journey is likely affected by more factors, such as those related to various aspects of the journey (not only the delay), hence the need to control for these journey aspects. Ideally, when modelling overall satisfaction, one would prefer using exogenous explanatory variables, i.e. not reported satisfaction with a given journey aspect. Due to the nature of the dataset, the overall satisfaction models include some explanatory variables that are endogenous (e.g., satisfaction with a specific aspect). This is because there are no measurable variables available that could serve as valid instruments for approaches such as instrumental variable estimation that could be used to address endogeneity. Consequently, these models may be subject to endogeneity bias.

It is highlighted that the reported values of R^2 are slightly higher than for the models of perception and delay satisfaction and, as previously noted, more similar to the values reported by Monsuur et al. (2021) in modelling overall journey satisfaction using NRPS data, what may be a result of the different nature of delay and overall journey satisfaction.

Using the estimated coefficients, lateness multipliers for arrival and departure delay were calculated for models 2-4 as shown in Table 58. The estimated lateness multipliers at arrival are around 4.0-4.7 for business travellers, 7.4-8.9 for commuters and 4.6-5.6 for leisure travellers. The respective departure lateness multipliers are 5.6-6.0 for business travellers, 2.1-3.5 for commuters and 3.2-4.0 for leisure travellers. The three estimated models in Table 58 indicate that the model results are quite robust to reducing the sample size or inclusion of additional control variables (as the estimated lateness multipliers are of similar magnitude for models 2-4). The lateness multipliers are larger at departure for business travellers, slightly larger at arrival for leisure travellers and much larger at arrival for commuters. This would suggest that 1 minute of delay is valued as being equivalent to around 4 minutes of scheduled journey time for delay at arrival and 6 minutes at departure for business travellers, 8 minutes at arrival and 3 at departure for commuters, and 5 minutes at arrival and 3 minutes at departure for leisure travellers. However, it is worth noting that while the central values are different, the estimated confidence intervals generally suggest that the lateness multipliers are not significantly different from each other. Nevertheless, it is not known how the multipliers reported in the literature performed with this respect, given that confidence intervals are typically not reported (i.e. Batley and Ibáñez., 2012).

With lateness multipliers being ratios of two values - utility of scheduled journey time and delay, the values depend on their relative magnitudes. At the same time, the observed differences in the scheduled journey times are typically relatively small (i.e. average absolute difference in timetabled scheduled journey times was just above 4 minutes for the OD pairs in the subsample used in model 4). This highlights the limitation of satisfaction data as (unlike in SP experiments), the analyst has no control over the attributes of the presented choice sets (or scheduled journey times and experienced lateness in the case of satisfaction surveys). It is possible that the small differences may remain unperceived by some passengers (perhaps even by frequent travellers) or have a lower marginal valuation as compared to larger differences. Daly et al. (2014) discussed the possible non-linearities in relation to time losses and savings whilst Wardman and Batley (2022) talked about the importance of time perception in relation to delays. If this is the case, the utility of scheduled journey time may be underestimated. In such cases, overcoming this limitation may be difficult as satisfaction scores are related to an *ex-post* evaluation of experiences and large changes in scheduled journey times are rarely observable.

Table 57 Modelling results for the initial models

	1	t-stat	2	t-stat	3	t-stat	4	t-stat
Constant								
C	-0.0778	-0.93	-0.225	-1.85	-0.186	-1.25	-0.117	-0.69
L	-0.0124	-0.14	-0.0553	-0.48	-0.0223	-0.16	-0.0793	-0.49
Station_Sat								
B	1.345***	20.3	1.350***	17.3	1.325***	15.1	1.137***	12.3
C	1.111***	27.6	1.183***	20.9	1.195***	16.4	0.980***	12.9
L	1.401***	27.9	1.465***	21.5	1.413***	17.1	1.172***	13.6
Train_Sat								
B	3.059***	45.8	3.066***	38.2	3.127***	34.0	2.866***	28.9
C	3.049***	72.2	2.998***	52.0	3.042***	41.6	2.770***	36.0
L	3.438***	62.2	3.418***	46.0	3.401***	36.6	2.999***	30.1
Freq_Sat								
B							0.803***	8.15
C							0.888***	12.3
L							0.919***	9.88
VfM_Sat								
B							1.049***	15.9
C							1.120***	15.0
L							1.123***	19.6
L_ A (β_3)								
B	-0.0505***	-8.82	-0.0567***	-8.89	-0.0521***	-7.68	-0.0537***	-7.53
C	-0.101***	-18.4	-0.1000***	-14.1	-0.114***	-13.1	-0.109***	-12.0
L	-0.0593***	-12.3	-0.0583***	-9.74	-0.0570***	-8.45	-0.0576***	-8.20
L_ D (β_2)								
B	-0.0690***	-7.77	-0.0683***	-6.46	-0.0729***	-5.99	-0.0758***	-6.01
C	-0.0522***	-7.71	-0.0472***	-5.20	-0.0296**	-2.60	-0.0349**	-2.97
L	-0.0404***	-6.30	-0.0421***	-5.02	-0.0354***	-3.54	-0.0402***	-3.84
SJT (β_1)								
B	-0.0009*	-1.99	-0.0120***	-5.43	-0.0121***	-5.00	-0.0134***	-5.27
C	-0.0058***	-6.76	-0.0134***	-5.26	-0.0142***	-4.74	-0.0123***	-3.93
L	0.0002	0.53	-0.0105***	-4.95	-0.0102***	-4.34	-0.0125***	-5.06
Thresholds								
1	-2.482***	-31.6	-2.890***	-4.55	-3.755***	-7.41	-3.501***	-6.47
2	-0.919***	-12.7	-1.243*	-1.96	-2.112***	-4.19	-1.815***	-3.37
3	0.772***	10.7	0.455	0.72	-0.410	-0.81	-0.0466	-0.09

4	4.396***	56.8	4.190***	6.60	3.293***	6.53	3.870***	7.17
Number of responses								
	40363		25457		17316		16632	
Log-likelihood								
	-36770.8		-22388.3		-15181.8		-13920.9	
Pseudo R^2								
	0.234		0.246		0.231		0.267	
Fixed effects								
	X		V		V		V	
VfM and Freq Satisfaction								
	X		X		X		V	
Minimum N								
	1		10		25		25	

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
B – Business, C – Commute, L – Leisure; Station_Sat, Train_Sat, Freq_Sat, VfM_Sat refer to satisfaction with station, train, frequency and value for money; L_A, L_D refer to delay at arrival and departure; SJT – scheduled journey time

Table 58 Estimated lateness multipliers

Journey Purpose	w_A (2)	w_A (3)	w_A (4)	w_D (2)	w_D (3)	w_D (4)
Business	4.74***	4.31***	3.99***	5.72***	6.02***	5.64***
z-stat	(4.67)	(4.22)	(4.34)	(4.12)	(3.81)	(3.94)
95% CI	[2.8-6.7]	[2.3-6.3]	[2.2-5.8]	[3.0-8.4]	[2.9-9.1]	[2.8-8.4]
Commute	7.43***	8.02***	8.86***	3.51***	2.08***	2.83***
z-stat	(4.90)	(4.44)	(3.73)	(3.75)	(2.31)	(2.41)
95% CI	[4.5-10.4]	[4.5-11.6]	[4.2-13.5]	[1.7-5.3]	[0.3-3.9]	[0.5-5.1]
Leisure	5.52***	5.61***	4.61***	3.99***	3.49***	3.21***
z-stat	(4.51)	(3.96)	(4.42)	(3.44)	(2.67)	(2.99)
95% CI	[3.1-7.9]	[2.8-8.4]	[2.6-6.6]	[1.7-6.3]	[0.9-6.0]	[1.1-5.3]

Legend: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The extended models

To enable closer comparisons with the work conducted by Batley and Ibáñez (2012) and provide sensitivity analysis, the ordered logit models were re-estimated using the extended version of journey type segmentation (presented in Table 54) with the results shown in Table 59. As in the case of the simpler models, model 1a is estimated without OD fixed effects that are subsequently included in models 2a-4a. Similarly, in model 1a, the scheduled journey time coefficient is negative and significant for commuters. With the inclusion of OD fixed effects in models 2a-4a, the scheduled journey time coefficient becomes negative and significant for all segments. At the same time, coefficients for arrival delay are negative and significant for all journey length and purpose combinations.

However, in the case of departure delay, despite all of the coefficients being negative, they are not significant for short commute and leisure. This may be a result of the additional segmentation leading to fewer responses by demand segment.

Table 59 Modelling results with extended segmentation

	1a	t-stat	2a	t-stat	3a	t-stat	4a	t-stat
Constant								
SC	0.0358	0.24	-0.0185	-0.09	0.0268	0.10	0.00342	0.01
SL	0.286	1.76	0.322	1.40	0.464	1.65	0.240	0.75
LB	0.0544	0.32	0.141	0.53	0.226	0.72	0.0669	0.19
LC	0.0142	0.08	0.0390	0.15	0.193	0.61	0.223	0.63
LL	-0.0295	-0.18	0.0242	0.09	0.183	0.59	-0.0142	-0.04
Station_Sat								
SB	1.718***	13.1	1.818***	10.5	1.823***	8.98	1.593***	7.33
SC	1.090***	23.0	1.141***	16.5	1.077***	11.53	0.864***	8.90
SL	1.376***	17.1	1.427***	11.7	1.264***	7.82	0.955***	5.70
LB	1.168***	13.0	1.127***	11.4	1.053***	9.36	0.899***	7.54
LC	1.126***	12.1	1.118***	9.56	1.300***	9.20	1.098***	7.44
LL	1.340***	18.2	1.386***	15.3	1.360***	12.65	1.169***	10.4
Train_Sat								
SB	2.845***	23.0	2.826***	17.7	2.870***	15.5	2.629***	12.9
SC	3.079***	63.9	3.043***	44.2	3.158***	34.91	2.863***	29.8
SL	3.286***	41.1	3.290***	28.2	3.369***	22.0	3.006***	18.3
LB	3.119***	36.6	3.159***	31.6	3.260***	28.4	3.024***	24.2
LC	2.992***	34.8	2.941***	27.3	2.808***	21.5	2.574***	18.4
LL	3.666***	46.7	3.625***	36.0	3.566***	28.8	3.110***	22.9
Freq_Sat								
SB							0.734***	3.46
SC							0.895***	9.89
SL							1.096***	6.72
LB							0.688***	5.28
LC							0.731***	5.05
LL							0.794***	6.19
VfM_Sat								
SB							0.884***	6.12
SC							1.065***	11.4
SL							0.972***	9.71

LB							1.114***	12.8
LC							1.215***	8.14
LL							1.219***	15.53
L_ A (β_3)								
SB	-0.0892***	-6.17	-0.109***	-6.50	-0.102***	-5.72	-0.106***	-5.64
SC	-0.101***	-15.4	-0.0978***	-11.1	-0.114***	-9.94	-0.106***	-8.83
SL	-0.0658***	-7.01	-0.0555***	-4.14	-0.0722***	-4.55	-0.0792***	-4.82
LB	-0.0429***	-6.51	-0.0473***	-6.52	-0.0422***	-5.45	-0.0463***	-5.66
LC	-0.109***	-9.45	-0.113***	-8.34	-0.123***	-8.07	-0.119***	-7.50
LL	-0.0576***	-9.77	-0.0623***	-8.93	-0.0574***	-7.40	-0.0545***	-6.71
L_D (β_2)								
SB	-0.0487*	-2.56	-0.0423	-1.82	-0.0578*	-2.19	-0.0627*	-2.28
SC	-0.0659***	-7.97	-0.0508***	-4.31	-0.0181	-1.12	-0.0295	-1.77
SL	-0.0560***	-4.78	-0.0633***	-3.77	-0.0252	-1.22	-0.0175	-0.80
LB	-0.0585***	-5.43	-0.0578***	-4.58	-0.0552***	-3.74	-0.0560***	-3.66
LC	-0.0261	-1.92	-0.0367*	-2.29	-0.0369*	-2.05	-0.0385*	-2.08
LL	-0.0332***	-4.12	-0.0345***	-3.40	-0.0362**	-3.01	-0.0454***	-3.64
SJT (β_1)								
SB	-0.00102	-0.70	-0.0115***	-3.44	-0.0109**	-2.81	-0.0115**	-2.84
SC	-0.00695***	-6.50	-0.0154***	-4.68	-0.0148***	-3.70	-0.0134**	-3.21
SL	-0.000697	-0.80	-0.0117***	-4.18	-0.0118***	-3.70	-0.0144***	-4.32
LB	0.000002	0.00	-0.0102***	-4.09	-0.0101***	-3.59	-0.0116***	-3.91
LC	-0.00334*	-2.00	-0.0114***	-3.49	-0.0127**	-3.28	-0.0121**	-2.97
LL	0.000119	0.31	-0.00960***	-3.99	-0.00906***	-3.35	-0.0115***	-4.02
Threshold 1	-2.421***	-16.9	-2.779***	-4.18	-3.350***	-5.90	-3.298***	-5.41
Threshold 2	-0.842***	-6.03	-1.107	-1.67	-1.672**	-2.96	-1.581**	-2.61
Threshold 3	0.865***	6.21	0.595	0.90	0.0369	0.07	0.184	0.30
Threshold 4	4.499***	31.5	4.348***	6.55	3.755***	6.63	4.114***	6.76
N	36220		22397		14733		14146	
LL	-33007.8		-19694.8		-12928.6		-11862.8	
R^2	0.237		0.251		0.236		0.271	
Fixed effects	X		V		V		V	
VfM and Freq	X		X		X		V	
Minimum N	1		10		25		25	

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
SB/LB – Short/Long Business, SC/LC – Short/Long Commute, SL/LL – Short/Long Leisure;
Station_Sat, Train_Sat, Freq_Sat, VfM_Sat refer to satisfaction with station, train, frequency and value for money; L_A, L_D refer to delay at arrival and departure; SJT – scheduled journey time

The subsequently estimated lateness multipliers are shown in Table 60. Arrival delay multipliers are around 9 for short and 4 for long business journeys, 7 for short and 9 for long commute and 5-6 for leisure trips. The corresponding departure delay multipliers are 4-5 for business journeys, 1-3 for commute and 1-5 for leisure trips. There is less confidence in the departure delay multipliers as the estimated values tend to vary considerably between the estimated models.

Table 60 Estimated lateness multipliers for models with additional journey type categorisation

Journey Purpose	w_A			w_D		
	(2)	(3)	(4)	(2)	(3)	(4)
SB	9.46***	9.38***	9.19***	3.68	5.31	5.45
z-stat	(2.97)	(2.47)	(2.48)	(1.64)	(1.75)	(1.81)
95% CI	[3.2-16]	[1.9-17]	[1.9-16]	[-0.8-8.1]	[-0.6-11]	[-0.4-11]
LB	4.63***	4.20***	4.00***	5.66***	5.49***	4.84***
z-stat	(3.50)	(3.03)	(3.25)	(3.03)	(2.58)	(2.67)
95% CI	[2.0-7.2]	[1.5-6.9]	[1.6-6.4]	[2.0-9.3]	[1.3-9.6]	[1.3-8.4]
SC	6.33***	7.72***	7.91***	3.29***	1.22	2.21
z-stat	(4.29)	(3.48)	(3.02)	(3.22)	(1.08)	(1.56)
95% CI	[2.3-9.2]	[3.4-12]	[2.8-13]	[1.3-5.3]	[-1.0-3.4]	[-0.6-5.0]
LC	9.84***	9.64***	9.80***	3.20***	2.90	3.18
z-stat	(3.15)	(2.98)	(2.71)	(1.99)	(1.82)	(1.78)
95% CI	[3.7-16]	[3.3-16]	[2.7-17]	[0.1-6.4]	[-0.2-6.0]	[-0.3-6.7]
SL	4.76***	6.11***	5.49***	5.43***	2.13	1.21
z-stat	(2.94)	(2.89)	(3.23)	(2.80)	(1.15)	(0.79)
95% CI	[1.6-7.9]	[2.0-10]	[2.2-8.9]	[1.6-9.2]	[-1.5-5.7]	[-1.8-4.2]
LL	6.49***	6.34***	4.74***	3.60***	3.99***	3.95***
z-stat	(3.70)	(3.11)	(3.52)	(2.53)	(2.18)	(2.63)
95% CI	[3.1-9.9]	[2.3-10]	[2.1-7.4]	[0.8-6.4]	[0.4-7.6]	[1.0-6.9]

Legend: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sensitivity analysis

To test the sensitivity of the approach used in estimating the lateness multipliers, the models from Table 59 were also re-run with the exclusion of departure delay from the list of explanatory variables. The results of the estimated models are presented in Table 61. The estimated coefficients are (similarly to the previously estimated models) negative and significant for arrival delay and scheduled journey time when OD pair fixed effects are included while the coefficients for scheduled journey time are mostly insignificant in the case of model 1 where OD pair fixed effects are excluded. The estimated lateness multipliers for arrival delay are shown in Table 62. These are similar to the ones from the main body of analysis - around 10 minutes for short and 4 minutes for long business journeys, 7 minutes for short and 10 minutes for long commute and 5-6 minutes for leisure journeys.

Table 61 Model estimates for models without departure delay

	1	t-stat	2	t-stat	3	t-stat	4	t-stat
Constant								
SC	0.0268	0.18	-0.0156	-0.07	0.0693	0.27	0.0233	0.08
SL	0.292	1.82	0.320	1.41	0.506	1.81	0.270	0.85
LB	0.0122	0.07	0.0727	0.28	0.170	0.54	0.000517	0.00
LC	0.0404	0.23	0.0447	0.17	0.220	0.70	0.250	0.71
LL	-0.0248	-0.15	0.00372	0.01	0.171	0.56	-0.0463	-0.13
Station_Sat								
SB	1.727^{***}	13.23	1.824^{***}	10.59	1.835^{***}	9.05	1.597^{***}	7.35
SC	1.090^{***}	23.32	1.138^{***}	16.42	1.075^{***}	11.52	0.865^{***}	8.92
SL	1.383^{***}	17.81	1.432^{***}	11.75	1.258^{***}	7.79	0.950^{***}	5.68
LB	1.170^{***}	13.63	1.134^{***}	11.42	1.056^{***}	9.38	0.901^{***}	7.56
LC	1.120^{***}	12.05	1.117^{***}	9.58	1.297^{***}	9.19	1.103^{***}	7.49
LL	1.339^{***}	18.15	1.386^{***}	15.30	1.360^{***}	12.65	1.167^{***}	10.37
Train_Sat								
SB	2.848^{***}	23.01	2.831^{***}	17.69	2.871^{***}	15.52	2.631^{***}	12.89
SC	3.078^{***}	63.92	3.041^{***}	44.27	3.153^{***}	34.92	2.863^{***}	29.81
SL	3.282^{***}	41.11	3.284^{***}	28.19	3.369^{***}	22.07	3.003^{***}	18.32
LB	3.125^{***}	36.75	3.160^{***}	31.68	3.275^{***}	28.54	3.042^{***}	24.35
LC	2.993^{***}	35.00	2.947^{***}	27.51	2.815^{***}	21.66	2.581^{***}	18.51
LL	3.671^{***}	46.90	3.631^{***}	36.15	3.579^{***}	28.94	3.126^{***}	23.06
Freq_Sat								
SB							0.716^{***}	3.37
SC							0.887^{***}	9.81
SL							1.097^{***}	6.73
LB							0.684^{***}	5.25
LC							0.709^{***}	4.91
LL							0.792^{***}	6.17
VfM_Sat								
SB							0.889^{***}	6.16
SC							1.061^{***}	11.31
SL							0.972^{***}	9.71
LB							1.112^{***}	12.79
LC							1.224^{***}	8.20
LL							1.216^{***}	15.49
L_ A ($\boldsymbol{\beta_3}$)								

SB	-0.113***	-10.45	-0.127***	-9.73	-0.125***	-8.66	-0.130***	-8.57
SC	-0.139***	-29.95	-0.122***	-18.14	-0.122***	-13.58	-0.119***	-12.78
SL	-0.0954***	-13.50	-0.0896***	-8.97	-0.0847***	-6.95	-0.0877***	-6.95
LB	-0.0599***	-10.30	-0.0620***	-9.49	-0.0530***	-7.36	-0.0577***	-7.67
LC	-0.125***	-16.83	-0.135***	-15.15	-0.145***	-13.90	-0.142***	-13.15
LL	-0.0713***	-14.60	-0.0748***	-12.65	-0.0694***	-10.35	-0.0698***	-10.06
SJT (β_1)								
SB	-0.000484	-0.33	-0.0134***	-4.02	-0.0119**	-3.09	-0.0126**	-3.13
SC	-0.00545***	-5.19	-0.0168***	-5.11	-0.0159***	-4.00	-0.0145***	-3.49
SL	-0.000195	-0.22	-0.0132***	-4.74	-0.0130***	-4.08	-0.0158***	-4.74
LB	0.000552	0.97	-0.0117***	-4.69	-0.0112***	-4.01	-0.0130***	-4.41
LC	-0.00281	-1.71	-0.0129***	-3.95	-0.0137***	-3.57	-0.0134***	-3.32
LL	0.000382	1.01	-0.0113***	-4.71	-0.0104***	-3.85	-0.0130***	-4.56
Threshold 1	-2.348***	-16.65	-2.901***	-4.34	-3.552***	-6.28	-3.557***	-5.86
Threshold 2	-0.780***	-5.66	-1.239	-1.86	-1.881***	-3.34	-1.850**	-3.06
Threshold 3	0.919***	6.69	0.454	0.68	-0.178	-0.32	-0.0926	-0.15
Threshold 4	4.541***	32.26	4.198***	6.29	3.532***	6.26	3.830***	6.32
N	36220		22397		14733		14146	
LL	-33079.4		-19730.8		-12945.6		-11882.3	
R^2	0.235		0.250		0.235		0.270	
Fixed effects	X		V		V		V	
VfM and Freq	X		X		X		X	
Minimum N	1		10		25		25	

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; SB/LB – Short/Long Business, SC/LC – Short/Long Commute, SL/LL – Short/Long Leisure; Station_Sat, Train_Sat, Freq_Sat, VfM_Sat refer to satisfaction with station, train, frequency and value for money; L_A, refers to delay at arrival; SJT – scheduled journey time

Table 62 Estimated lateness multipliers for models without departure delay

Journey Purpose	w_A		
	(2)	**(3)**	**(4)**
Short Business	9.45***	10.47***	10.31***
z-stat	(3.67)	(2.87)	(2.90)
95% CI	[4.4-14.5]	[3.3-17.6]	[3.3-17.3]
Long Business	5.32***	4.74***	4.45***
z-stat	(4.24)	(3.56)	(3.86)
95% CI	[2.9-7.8]	[2.1-7.3]	[2.2-6.7]
Short Commute	7.28***	7.65***	8.19***
z-stat	(4.92)	(3.86)	(3.38)
95% CI	[4.4-10.2]	[3.8-11.5]	[3.4-12.9]
Long Commute	10.48***	10.54***	10.57***
z-stat	(3.83)	(3.47)	(3.23)

95% CI	[5.1-15.8]	[4.6-16.5]	[4.2-17.0]
Short Leisure	6.78***	6.50***	5.55***
z-stat	(4.18)	(3.52)	(3.93)
95% CI	[3.6-10.0]	[2.9-10.1]	[2.8-8.3]
Long Leisure	6.62***	6.68***	5.37***
z-stat	(4.44)	(3.63)	(4.19)
95% CI	[3.7-9.5]	[3.1-10.3]	[2.9-7.9]

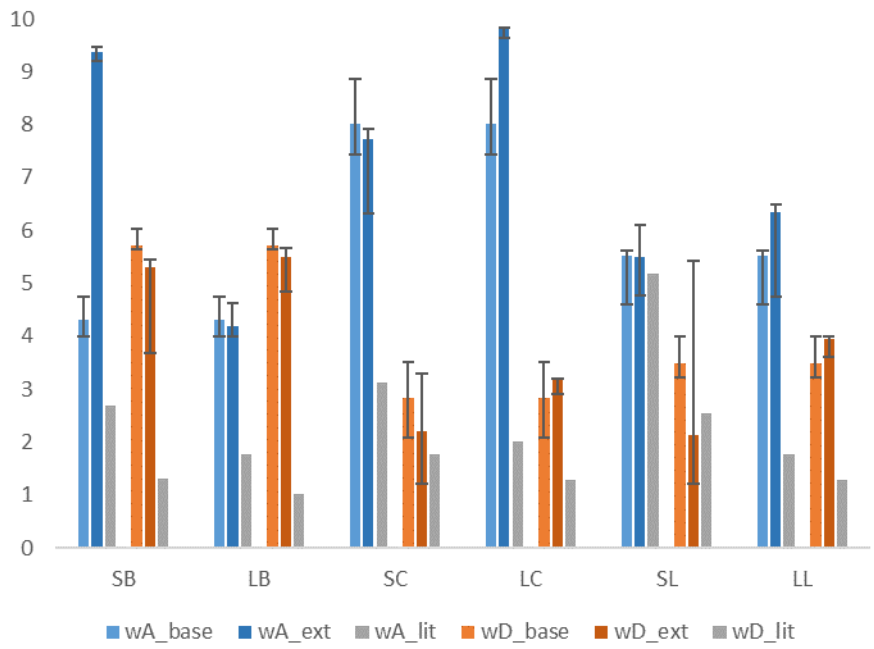
Legend: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8.5. Comparison of the estimated lateness multipliers with Batley and Ibáñez (2012)

The lateness multipliers from the initial (with three demand segments) and extended model (with segmentation by journey purpose and length conforming to the segmentation adopted by Batley and Ibáñez, 2012) were subsequently compared with the equivalent values estimated by Batley and Ibáñez (2012). These have been retrieved from Table 4b in Batley and Ibáñez (2012) for lateness multipliers at arrival whilst the lateness multipliers at departure were calculated from the estimated coefficients reported in Table 4a in Batley and Ibáñez (2012). These are presented in Figure 60.

For the initial models where no journey type categorisation was introduced, arrival lateness multipliers of respectively 4, 8 and 5 were estimated for business, commute and leisure travellers. With the introduction of journey type categorisation, the estimated multipliers for business travel were suggested to be larger for short journeys (around 10) and comparable to the initial model for long journeys (around 4). For commuter and leisure, comparable lateness multipliers were estimated from the initial and extended models - between 7 and 10 for commuters (slightly larger for the long journeys) and around 4-6 for leisure.

In the case of departure lateness multipliers, the values estimated from the initial and extended models are very similar in all cases and do not seem to differ between short and long journeys. These are around 5-6 for business journeys, 2-3 for commute and 3-4 for leisure journeys.



Legend: wA denotes lateness multipliers at arrival, wD - at departure; 'base' refers to the model with 3 demand segments, reported in Table 58, 'ext' refers to the models with additional demand segmentation reported in Table 60; 'lit' refers to the equivalent values estimated by Batley and Ibáñez (2012); for each of the estimated multipliers the error bars refer to minimum and maximum estimated value from models 2-4, while the median value is reported as the central value.

Figure 60 Comparison of estimated lateness multipliers for delay at arrival

The results are indicative of lateness valuation increasing with journey time for commuters, decreasing for business travellers and being constant for leisure travellers. Hence, business travellers are suggested to be less concerned with delays relative to scheduled journey time for longer journeys – this may result from the productive use of in-vehicle times (as discussed in Wardman and Lyons, 2016; Lyons et al., 2016) and/or correlation between journey length and quality. However, in the case of commuters, the importance of delays relative to journey time seems to increase slightly with journey length. This may be related to the general dissatisfaction with long commute as noted in Cats et al. (2015) or correlations with other journey aspects. Though, it is worth highlighting that the lateness multipliers estimated here are not significantly different from each other. However, noting that confidence intervals are not generally reported in the literature for lateness multipliers, what makes detailed comparisons difficult.

Nevertheless, the estimated lateness multipliers seem to be larger than the values estimated by Batley and Ibáñez (2012). This can be possibly related to the different nature of satisfaction data or a relatively less negative disutility of scheduled journey time estimated from the satisfaction models due to the aforementioned data limitations. In line with Batley and Ibáñez (2012), in almost all cases (apart from long business journeys), the arrival delay multiplier is larger than the departure delay equivalent, indicative of the final (destination) delay being typically a source of larger disutility. However, it has to be noted that there are differences in the flows used for the analysis and some of the differences in the estimates can also be contributed to the differences in the journey type segmentation methods.

8.6. Conclusions

This work adds a degree of novelty in using passenger satisfaction data instead of the typically employed SP survey data to estimate lateness multipliers, a conversion rate between the value of a minute of lateness to the equivalent length of scheduled journey time. This combined the previous work using life satisfaction surveys in economic valuation (i.e. Layard et al., 2008) with work using passenger satisfaction surveys to study the impact of delays on passengers (i.e. Monsuur et al., 2021) and studies using SP surveys to estimate lateness multipliers (i.e. Batley and Ibáñez, 2012). A subset of the NRPS dataset provided by Transport Focus was used to create a pseudo-panel of frequent rail travellers to estimate an ordered logit model of passenger satisfaction with origin-destination pair fixed-effects to estimate the utilities of delay and scheduled journey time. Subsequently, their ratios were calculated to derive the lateness multipliers.

The estimated lateness multipliers are slightly larger than the ones typically estimated in the SP studies and some caution is needed while applying these values. To the best of knowledge, it is the first study attempting to use journey satisfaction data in such an application. It does, however, highlight the potential of using such data in transport economics as the estimated coefficients and resulting multipliers are of expected signs and magnitudes.

The results suggest that:

1) In most cases, delay at arrival is a source of larger disutility relative to delay at departure. This finding is in line with expectations and consistent with Batley and Ibáñez (2012). However, the opposite is suggested to be true for long business journeys, possibly due to the ability to more productively use the additional in-vehicle time related to on-board delays (for discussion on productive use of travel time see Wardman and Lyons, 2016; Lyons et al., 2016).
2) Relative to scheduled journey time, 1 minute of delay at arrival is valued as an equivalent of around 9 minutes of SJT for short business and long commute, 7 minutes for short commute, 5 minutes for leisure and 4 minutes for long business

journeys. These values are typically larger than the estimates from SP studies with a minute of lateness valued up to 3 times more relative to scheduled journey time as compared with Batley and Ibáñez (2012). While the comparisons presented above are made in relation to the estimated values from Batley and Ibáñez (2012), it is noted that Wardman and Batley (2014) reported lateness multipliers between 2-5 for business travellers and commuters and 2-7 for leisure travellers with even larger values being reported in Wardman (2001) and Börjesson and Eliasson (2011).

3) The valuation of lateness at arrival with respect to scheduled journey time is suggested to decrease with journey length for business travellers, increase for commuters and remain constant for leisure travellers. This is slightly different to the results obtained by Batley and Ibáñez (2012), suggesting that the valuation decreases with journey length.
4) Relative to scheduled journey time, 1 minute of delay at departure is valued at around 5 minutes for business journeys and 2-4 for other travellers. While the differences are not necessarily statistically significant, the larger disutility related to delay at departure for business users may be due to the impact of such a delay on their productive use of travel time, which is of paramount importance for such travellers. These values are also larger than the estimates from SP studies with a minute of lateness valued up to five times more relative to scheduled journey time as compared to Batley and Ibáñez (2012).

The analysis presented as part of this chapter highlights the potential of satisfaction data in economic valuation. As noted previously – it has been previously applied in economic valuation in health, labour and environmental economics. However, application in transport has been very limited. Comparing satisfaction data to SP, RP or ticket sales data can offer some additional insights regarding the negative impacts (i.e. of delays) that are not reflected in hypothetical (i.e. SP) or actual (i.e. RP or ticket sales) choices. As noted by Batley et al. (2011) the higher valuation of lateness at the individual level, but a much more limited impact of delays on demand may be explained by delays having a very negative impact on passenger satisfaction, but not necessarily leading to changes in travel choices. If worsening performance does not lead to changes in behaviour (choices), these still have an impact on social welfare. As noted by numerous articles, there is an increasing need to look at alternative ways of measuring social welfare (e.g. Fleurbaey, 2009). Therefore, using journey or life satisfaction data and relating it to the supply of public transport as well as its performance can become a valuable addition to the standard economic approaches in transport appraisal.

It is believed that applying the methodology used in this study to similar journey satisfaction datasets would be useful in further exploring the potential of such data sources. In doing so, it is worth considering the following limitations of the NRPS dataset:

1) The dataset is cross-sectional in nature. The presented approach uses numerous assumptions to construct a pseudo-panel of frequent travellers. Using a true panel dataset would, therefore, be preferable. Moreover, the analysis of SP surveys typically focuses on the mean or standard deviation of performance (e.g. Batley and Ibáñez, 2012). However, the recorded delay lengths and satisfaction scores in the NRPS survey refer to incidental journey experiences.
2) The observed changes in scheduled journey time as well as delay lengths are naturally beyond the researcher's control. With most changes in scheduled journey time being relatively small, it is possible that the utility of scheduled journey time is less negative than it would be for larger changes. This could be the reason behind the estimated multipliers being larger than the ones typically estimated in the literature. Therefore, conducting similar surveys may be useful for OD pairs that present interesting case studies, allowing for studying the impact of smaller and larger differences in journey times. Ensuring representation of delays of differing lengths is more difficult, as ideally, the surveys would need to be conducted over a long time period of time (as was the case with the NRPS) to increase the chances of observing shorter and longer delays.
3) While the results offer some useful insights into the valuation of delays, it is worth noting that the estimated lateness multipliers are generally not significantly different from each other, what may be a feature of using satisfaction data and the issues summarised above.
4) The NRPS dataset may also be prone to data errors related to travel records as in some cases passenger reports of delay experiences were significantly different from the recorded performance, especially for the longer delays. This may be due to possible differences between the journeys travellers planned to make, actual, reported journeys and interchanges (as described in the previous chapters).

Whilst for future studies, it is recommended that the aforementioned limitations are considered, the approach presented in this chapter led to the estimation of the utility of scheduled journey time and lateness that were both of expected direction (i.e. negative) and magnitude (i.e. delay coefficient being more negative than that of scheduled journey time), which shows the potential that satisfaction data has in economic valuation.

The approach used here was based on demand segmentation aligning with the work by Batley and Ibáñez (2012). However, some alternative segmentations could be applied based on sociodemographic characteristics. Moreover, a cluster analysis could be applied to identify the different types of passengers from the dataset rather than using the pre-defined categorisation of respondents.

It is recommended that accurate data on journey history is collected from the satisfaction surveys to allow detailed investigation of the journey that traveller was planning to make and their actual experience to reduce the scope for errors. Moreover, including more questions related to income and fares in the questionnaires could allow the estimation of more sophisticated metrics and limit the potential for endogeneity bias related to using satisfaction variables as explanatory variables. In the case of NRPS, the possible lines of investigation include looking at the relationship between income, fare, scheduled journey time, headway, performance, journey quality and satisfaction with value for money, possibly even allowing calculation of the value of time.

Chapter 9

Conclusions

9.1. Introduction

Motivated by previous research and limited understanding of the impacts of delays on travellers, this thesis aimed to investigate rail passengers' perception of delays, impacts of delays on journey satisfaction and delay valuation. The thesis explored the intermediate steps linking the occurrence of delays and their impacts on demand and revenues, with a particular focus on understanding the impacts of different lengths of delays on passengers, given that smaller measured delays may not be noticed by passengers. The focus was on British railways given it is a well-developed case study in railway planning and with rich data on performance, finances and passenger satisfaction. The research can be especially helpful in designing/rethinking passenger delay compensation (reviewed as part of this thesis) and devising performance strategies and targets for railways (particularly in the British context). The aim was achieved by:

1) reviewing the current rail passenger delay compensation scheme rules, its impacts on passengers and operators, highlighting the current issues with the delay compensation scheme and advising future research directions as well as the considerations for the design of such a scheme in Chapter 4 (providing motivation for research conducted as part of the remaining chapters),
2) examining how passengers perceive delays in Chapter 6 to help understand whether some of the smaller delays remain unperceived and to devise the delay length perception thresholds,
3) investigating how recorded delays impact upon passenger satisfaction and any potential non-linearities in delay impacts in Chapter 7,
4) contrasting the concepts of delay perception and satisfaction to investigate whether the inability to perceive some of the smaller delays may explain the reaction of passenger satisfaction to measured delays in section 7.5. and
5) estimation of lateness multipliers (a conversion rate between a minute of delay and an equivalent length of scheduled journey time) using satisfaction instead of the typically utilised stated preference data in Chapter 8.

9.2. Summary of the key results

The work conducted as part of this thesis was motivated by the limited impact that performance has on rail demand. Travellers may be unable to change their travel behaviour in response to worsening performance due to limited availability of alternatives, what is reflected in relatively low values of estimated elasticities in the literature (ATOC, 2004;

Batley et al., 2011)., but this is not equivalent to delays having no impact on passengers. Though, it was noted that delayed travellers can claim compensation for the longer delays as part of the ‘Delay Repay’ scheme in Great Britain.

Against this background, the literature review conducted suggested that travellers are more likely to claim compensation for longer delays and if they paid more for their tickets, which is likely to be a result of how the scheme rules are constructed. As the scheme rules were introduced arbitrarily and are largely homogeneous, the first objective of the thesis was to evaluate the impact of this scheme on both passengers and the operators. While passengers may value the existence of a compensation scheme, it is difficult to estimate its benefits. These can be related to either increased demand due to the existence of the scheme (i.e. the scheme serving as delay insurance) or limited revenue losses (i.e. compensation serving as a way of retaining demand). The econometric analysis suggested that the homogeneity of scheme rules leads to an increased revenue burden for TOCs operating longer journeys, resulting from longer journeys being usually more likely to be affected by longer delays, what increases the proportion of passengers eligible to claim compensation. At the same time, with such journeys being typically more expensive, the expected compensation values are typically larger what encourages a higher proportion of the eligible passengers to submit claims. On average, each additional minute of APL was estimated to increase the proportion of ticket revenue repaid to passengers as part of the scheme by 0.2%. Additionally, due to increased engagement, for the same levels of performance, TOCs repay an additional 0.2% of their ticket revenue for each £10 of average fare. The increasing revenue impact of the scheme was suggested to be in line with some of the previous analysis of passenger engagement. It was, however, noted that the amount of research looking at the impact of delays on passengers is not sufficient to enable designing the scheme based on economic evidence. Rather, it was noted that to design a passenger delay compensation scheme, it is necessary to better understand what levels of delays are especially inconvenient for travellers. Hence, motivating further research and encouraging further interest from the industry and regulatory bodies.

First, the analysis of traveller perception suggested that passengers are highly unlikely to perceive delays of up to 2 minutes. It was estimated that commuters’ perception delay thresholds are between 2-8.5 minutes. For the other types of travellers, the respective delay lengths are between 3 and 20 minutes with the probability of perceiving a delay increasing with length of delay at both departure and arrival and decreasing with journey length and for seated passengers.

Regarding the ability to accurately perceive the lengths of experienced delays, a large proportion of reported lengths of delays is concentrated around multiples of 5s (almost 60% against 15% of recorded delays being of such lengths). However, a relatively large proportion of reported delay lengths is also scattered around non-multiples of 5s in the case of smaller delays of up to 8 minutes. This indicates on passengers who perceive the smaller delays being more likely to report the lengths of experienced delays more accurately.

The analysis of perception was suggested to be of limited application in economic appraisal, though, it can possibly provide an explanation behind the possibility for smaller delays having a more limited impact on passengers what was subsequently investigated.

In the case of satisfaction modelling it was highlighted that the binary representation of the satisfaction scale has merits over the original ordinal scale, being more amenable to policymaking. However, noting that any interpretation or classification of satisfaction versus dissatisfaction is generally open to interpretation and will affect the results. The results of the preferred delay satisfaction model suggested that in the case of incidental delays, commuters are unlikely to remain satisfied with punctuality if the length of delay is over 5-10 minutes. In the case of the other travellers, the respective lengths are in the range of 10-20 minutes. This is suggested to be slightly lower than the 30 minutes previously suggested by the literature. In the case of average performance, it was suggested that under perfect performance, 80% of commuters and 90% of other travellers are satisfied and each minute of APL leads to a reduction in the proportion of satisfied passengers by (on average) 6 pp for a minute of APL for commuters and 2 pp for other travellers. Typically, longer journeys and better journey quality can be attributed to lower probability of being dissatisfied with a delay of a given length whereas the probability of being dissatisfied with a delay typically increases if the service was also delayed at departure. This research also highlighted the possible non-linearities in the delay impacts on passengers that can be related to journey length (as the impact of delay on passenger satisfaction typically decreases with journey length) and smaller delays being generally less likely to be perceived and significantly affect satisfaction. The analysis aiming at estimating the elasticity of the marginal disutility of lateness led to inconclusive results, being largely affected by the choice of methodology. Nevertheless, the suggestion that some non-linearities may be present is in line with both the literature and also the perception, and satisfaction research conducted as part of this thesis. The proposed models of delay satisfaction can be used in setting performance targets or designing compensatory mechanisms.

Contrasting the concepts of delay perception and satisfaction suggested the existence of a gap between the two concepts as the estimated probability of perceiving a delay is larger than the probability of being dissatisfied with a given delay incident. The estimated gap is suggested to be between 0 minutes for standing commuters to up to 9 minutes for seated leisure travellers. This suggested that for commuters perceiving a delay almost automatically translates to an impact on satisfaction. However, for the other travellers, it was suggested that the estimated gap is also affected by journey quality as on condition that travellers are seated, the gap between the perceived delay and one largely affecting satisfaction is larger.

With perception and satisfaction having a limited application in economic appraisal, the work was extended to estimating lateness multipliers (the valuation of lateness with respect to scheduled journey time) using satisfaction data. As typically SP studies are used in their estimation, it was thought that comparing the valuation estimated using the hypothetical choice data and satisfaction data (representing evaluation of actual experience), may provide additional insights into the impacts that delays have on travellers. A minute of delay at arrival was estimated to be valued as equivalent of 4-9 minutes of scheduled journey time whilst a minute of delay at departure was estimated to be valued as equivalent of 2-5 minutes of scheduled journey time. The estimated values are larger than the lateness multipliers from Batley and Ibáñez (2012). However, it was noted that there are instances where values of similar magnitude were reported throughout the literature. The valuation of lateness at arrival was also suggested to decrease with journey length for business travellers, increase for commuters and remain constant for leisure travellers.

Overall, the thesis highlighted the fact that small delays may often have a more limited impact on passengers, whilst also suggesting that journey purpose, length, comfort and delay at departure are all important in determining the impact that delays have on passengers. In most cases, however, the delays that are likely to cause dissatisfaction, are much smaller than the current compensation thresholds. Nevertheless, the current scheme has a larger financial impact on TOCs operating longer journeys whereas the impact of delays for travellers on longer journeys was suggested to be less significant. Hence, the following section will summarise how the results of the work conducted as part of the thesis can be applied in practice, highlighting the main policy implications.

9.3. Policy implications

This section aims to position the results of the research conducted as part of this thesis and set out some recommendations for policymakers. These are not the direct results of the empirical analysis, but rather recommendations and measures through which the results can potentially be applied when designing performance and compensation schemes and/or seeking to reduce the negative impact of delays on passengers.

The investigation of the currently used rail passenger delay compensation scheme revealed that long-distance train operators are likely to repay a larger proportion of their revenues to passengers under the current scheme. The literature review and analysis conducted suggested that:

- the delay length compensation thresholds have been set arbitrarily and there is a need to understand whether a different design of the scheme may be more optimal,
- automation of the current scheme would lead to a larger portion of revenue repaid to passengers which motivates the need for research into the benefits versus costs of the currently used scheme,
- there is a need to explain if there are some regulatory or administrative reasons behind the homogeneous design of the scheme that lead to such differences,
- there is a precedent, particularly in the Czechia and Spain, for the eligibility criteria to vary with types or lengths of journey, and such a design of compensatory mechanisms could also be considered for the British railways,
- while the one-fit-for-all approach that is currently used may have some underlying limitations, it is also noted that a homogeneous, easy-to-understand and run scheme may have some benefits, both for passengers (who can navigate the rules easier), and the operators (reducing administrative costs) and
- regardless of the changes in the scheme, it could be centrally operated, which would greatly improve the homogeneity in its operation and facilitate the claiming processes for passengers.

The concept of delay perception has not been very well studied before and while the results may be intuitive (i.e. the probability of perceiving a delay increases with delay length), an attempt was made to estimate the delay length perception thresholds, suggesting the lengths of delay where passengers become more likely to perceive late running (than not). It must, however, be noted that the application of the concept of delay perception is likely to be limited in economic appraisal, i.e.

- the policymakers or train operators should not be focusing on targeting delay perception (i.e. reducing the probability that travellers perceive delays) as such solutions are not likely to be welfare maximising,
- conversely, it is believed that providing real-time information and clear communication to travellers about any possible delays may increase passenger

satisfaction as it is likely to reduce the uncertainty and inconvenience caused by the delays,

- the fact of perceiving or not perceiving a delay is not immediately connected to experienced utility or choices,
- rather, perceiving or not perceiving a delay is likely to be a reason for some delays having a more limited impact on passenger satisfaction as for delays to have an impact on passengers, they would typically need to be perceived first.

The concept of delay satisfaction is one that can have more application in policymaking as it can possibly aid the process of setting performance targets or delay compensation rules where:

- the individual-level satisfaction models can help set performance targets related to the distribution of lengths of delay incidents as smaller delays are less likely to have a negative impact on passenger satisfaction,
- whereas the aggregated OD-level models can help design performance targets related to average performance.

Additionally, with large impacts of journey quality, length and length of delay at departure related to how the final (destination) delays affect passenger satisfaction, it is advisable to set performance targets:

- related to lengths of delay at departure (not only at arrival),
- related to crowding levels, especially for late running episodes,
- varying by journey lengths as the analysis conducted in this thesis provided additional evidence that the impact of delay varies with length of journey (in line with Wardman and Batley, 2022 suggesting that proportional elasticities better explain changes in demand) and also
- varying by type of route as in absolute terms, OD pairs with a large commuter focus are likely to suffer from lower levels of satisfaction (for the same level of performance).

Ultimately, the estimated gap between lengths of delays that are perceived and start having a detrimental impact on passenger satisfaction also highlights the impact of journey quality on how delays affect passenger satisfaction. Hence, it is advisable to:

- in the event of delays, ensure that passengers are well-informed about the revised departure and arrival times,
- implement ways of reducing crowding on delayed services as worse journey quality will further reduce passenger satisfaction and

- limit cancellations or truncations of delayed services if these could lead to increased crowding.

As the lateness multipliers based on satisfaction data estimated as part of this thesis were larger than the ones typically estimated from SP studies, this also suggests that delays might have a more negative valuation relative to scheduled journey time than previously thought. Nevertheless, the satisfaction data may be a source of additional information that may be used by policymakers to monitor the impacts of delays on passengers as:

- the demand response following delays has been suggested to be relatively modest in the literature,
- changes in satisfaction do not ultimately lead to a choice but are likely to have an impact on well-being,
- hence, there is an argument for including the analysis of the impact of delays on travellers' satisfaction and/or well-being and incorporate these into the economic appraisal,
- estimation of the traditionally used metrics using satisfaction data may be a useful approach that allows better understanding of how passengers value different aspects of journeys and provide some sensitivity testing overcoming some of the limitations of SP surveys.

While this section summarised the wider application of the analysis conducted as part of this thesis, the remaining sections will discuss the potential to improve and/or further develop the presented analysis.

9.4. Limitations

The investigation of passenger delay compensation scheme concluded that long-distance TOCs repay a larger proportion of their revenues as part of the scheme due to the differences in the nature of their operation (i.e. journey length correlated with delay length) as well as the increased claiming rates related to a larger average value of compensation per claim on such journeys. However, the analysis was only conducted using very aggregate data (i.e. annual data at TOC level). Moreover, it was noted that there might be some benefits related to having one set of rules for all passengers rather than providing a very complex compensation scheme that would be less easy to understand and more costly to administer. Hence, it was suggested that a large-scale study be conducted to estimate the benefits and costs of the DR scheme as well as review its current design.

Throughout the thesis, several issues have been mentioned that are related to the use of the NRPS dataset. While the impact of most of them is likely to be small and, as part of the analysis of delay perception, an extensive sensitivity analysis was presented with the aim of testing the impact of any possible limitations, it is worth noting that:

- the questionnaires are handed out prior to passengers boarding trains what can have an impact on their responses and, hence, survey results;
- passengers filling in the survey closer to the time of completing a journey may have a better recollection of the journey but equally any negative emotions may be stronger when a negative experience (e.g. a delay) is more recent;
- some discrepancies between reported and recorded delays have been found;
- for passengers with multiple interchanges, it is difficult to establish to what extent passengers truly score satisfaction with one leg of their journey or their whole experience;
- the delays described by the dataset are related to delay incidents, not average performance. Hence, in the cases where models of passenger satisfaction are estimated at the aggregated, OD level, the levels of passenger satisfaction, in fact, relate to incidents of lateness averaged across all the responses;
- the dataset is cross-sectional in nature with multiple assumptions used to create a pseudo-panel of frequent travellers and subsequently estimate an ordered logit model with OD fixed effects to derive lateness multipliers. However, using a true panel dataset would be preferable.

Related to the choice of methodologies, the binary representations of the presented satisfaction models are highly sensitive to merging the five original satisfaction categories into two. This approach assumes that the distances on the scale between the points merged into each of the two categories are insignificant. On the other hand, the ordered logit models utilising the original five satisfaction scores are more difficult to interpret, as it is difficult to discern what is implied by a change between satisfaction scores. Similarly, some of the linear representations of the satisfaction variables assume that the distances between the satisfaction scores on the scale are equal, allowing averaging of satisfaction scores. Moreover, due to the non-quantitative nature of satisfaction, any policy targets set based on this work, are highly sensitive to the assumed definition of satisfaction versus dissatisfaction and the limitations of the approach used.

Some of the proposed models did not include all levels of the interacted variables, aiming to capture the complementary effect of these variables on the effect of delay length on perception or satisfaction. When the fully specified models were estimated, these failed to capture the proposed relationship correctly. For example, the estimated probabilities of perceiving a delay were suggested to start decreasing with increasing levels of delay at arrival at certain levels of delay at departure. The original models' specification does not allow for such a relationship due to only accounting for the complementary nature of these explanatory variables.

It is also noted that some of the overall satisfaction models that included satisfaction with specific journey aspects as explanatory variables may suffer from endogeneity bias. This is due to the fact that where some journey aspects are not captured by any of the measurable variables in the survey, reported satisfaction with those journey aspects is used as a proxy, hence introducing possible endogeneity bias.

More generally, all data used in the analysis refers to pre-COVID times and several studies addressed the impact of the COVID-19 pandemic on travel behaviour (e.g. Coppola and De Fabiis, 2021; Vickerman, 2021; Hörcher et al., 2022; Bonera and Martinelli, 2023; Deole et al., 2023). It is possible that the impact of delays on travellers is different now as compared to the pre-COVID times.

Finally, it must also be noted that the focus of this thesis remained on GB railways and, application of the results to other countries or industries may not be directly possible. However, it is believed that the methodologies outlined in this thesis may be applied to other contexts too.

With this section summarising the main limitations of the work conducted as part of this thesis, the following section aims to highlight some key areas that would benefit from further research.

9.5. Further research

Related to the policymaking applications previously highlighted in section 9.3, it is recognised that there are several ways in which the analysis presented as part of this thesis could be enhanced or extended.

First of all, it is advisable that the regulators and policymakers conduct a more advanced analysis of the costs and benefits of the currently operating passenger delay compensation scheme (perhaps using data at the OD level) to identify:

- the value that the scheme has for passengers and the impacts of the scheme on passenger satisfaction, demand and operator revenues,
- whether the rules of the compensation scheme are optimal, possibly exploring the alternative ways of design, including a version of the scheme where the rules vary by service type or journey length,

- how changes in ticketing or automation of the scheme would impact upon the scheme operation and costs and
- the role and impacts of the compensation scheme before and after the COVID pandemic.

Secondly, with this thesis presenting an extensive analysis of delay perception and delay impacts on satisfaction, it is advisable that studies aiming at using similar datasets include:

- more detailed information about the journey plans versus the actual journey including specific information about all the interchanges to reduce the scope for errors and
- questions related to income and fares that, along the satisfaction with value for money (forming part of the NRPS), could allow estimation of the value of time or similar metrics.

The analysis of delay satisfaction poses a different question that relates to how to incorporate passenger satisfaction in economic appraisal. Several ways to pursue this direction could include investigating:

- the link between passenger satisfaction and demand, and
- the relationship between travel satisfaction and wellbeing.

Subsequently, the thesis highlighted the potential for non-linearity in the delay impacts on passengers related to both the shorter delays not being perceived (thus likely having a smaller marginal valuation) as well as the impact of delays typically decreasing with journey length. In most cases, both delay and scheduled journey times enter the estimated models as continuous variables. However, alternative ways of representing these variables could also be used, for example including the relative length of delay as compared to the scheduled journey time. Moreover, the investigation of the elasticity of the marginal disutility of lateness led to some inconclusive results, suggesting that some non-linearities are present. Hence, it is advisable that alternative datasets and/or methodologies are implemented to investigate such non-linearities as well as advising on ways of implementing them in the economic appraisal.

Finally, this thesis represents one of the very few attempts to use passenger satisfaction data in the economic valuation in transport. It highlights the potential of such data and future studies might want to repeat a similar exercise using a different satisfaction dataset that could help overcome some of the issues with the NRPS as well as numerous assumptions implemented. It is advisable to conduct such surveys as panel, not cross-sectional studies that could then ask for satisfaction with a particular journey and/or general satisfaction, and for selected case studies, i.e. looking at interesting OD pairs where travel alternatives vary with respect to journey times and prices. However, as noted previously, in the case of satisfaction surveys, it is difficult to investigate the impact of delays of varying lengths as these are beyond the researcher's control. Hence, such a study would need to be conducted over a relatively long time period to ensure that the distribution of delays represents both shorter and longer journeys. This could lead to an RP study where participants are also asked to score their satisfaction, noting that the proposed study would be relatively complex and expensive to implement.

Annex I

Sensitivity analysis for models of delay perception from Chapter 6

The purpose of this annex is to present the results of the sensitivity analysis conducted as part of the analysis of delay perception from Chapter 6. A number of sensitivity tests was conducted to understand the sensitivity of model results to the used methodology or data errors.

The sensitivity analysis includes:

1) Estimation of probit, instead of logit, models of delay perception to investigate the impact of the chosen methodology (i.e. logit versus probit) on the estimated results.
2) Estimation of model 1 only for the responses where no delay at arrival was recorded to investigate whether some of these travellers may have been biased by the existence of delay at departure (section 6.5). As such, the explanatory variable representing length of delay at arrival is replaced by the length of delay at departure.
3) Restricting the sample used in the main body of analysis only to leisure travellers with restricted tickets who are only able to travel on a specific service. Hence, in case of delays they cannot use a previous service that had been delayed or the following service in case of delays to their service. This was expected to improve the accuracy of matching passenger reports of journey and operational data. It has previously been noted (in section 5.2.2) that some discrepancies are expected and especially so in the case of longer delays. This could be a result of differences between the travel plans and actual journeys made. For example, passengers might have originally wanted to travel on one service but travelled on a different one instead. Similarly, in the cases where passengers interchanged, there might be confusion related to whether passengers' reports relate to this specific leg of the journey or the whole journey. Hence, the dataset was restricted to passengers on restricted tickets and with no interchanges.
4) Relaxing the assumptions from for the non-interchanging travellers (section B)
5) Ultimately, model 1 is re-estimated using an extended version of journey purpose segmentation (section C) where all the six previously introduced journey categories are interacted with additional seven journey type categories (i.e. airport, high-speed, interurban, long commute, long distance, rural, short commute). These replace the previously used variables describing the length of the journey as journey length and type are correlated. It is believed that these journey type categories provide some additional insights as these are proxies for some characteristics related to the specific journey types and may be sources of heterogeneity in perception.

This is summarised in Table 63:

Table 63 Summary of the sensitivity analysis

Sensitivity analysis	Section	Description	Impact	Sample size
1 - Probit	A	Estimation of probit, instead of logit models of delay perception	Very small; the estimated probabilities slightly lower for probit due to logistic distribution having longer tails (Horowitz and Savin, 2001).	48,793 to 73,050
2 – Restricted tickets with no interchanges	B	Logit model is run on a sample restricted to leisure travellers with advanced tickets and no interchanges	Differences between the restricted and full model are insignificant, though, the predicted delay perception thresholds are slightly lower.	1,241 to 3,281
3 – No interchanges	B	Logit model is run on a sample restricted to travellers with no interchanges	Results similar to the full model, however, some coefficients are insignificant (possibly due to smaller sample sizes).	9,075 to 29,125
4 –Additional journey type categorisation	C	Logit model in its initial form (i.e. model 1) is run with the addition of interaction between the length of delay and 7 journey types	Differences between arrival delay coefficients are typically insignificant, probably due to smaller sample sizes. However, the coefficient on arrival delay is generally larger for the journey types characterised by shorter journey lengths (i.e. probability of delay perception increases more rapidly), in line with the previously estimated models.	73,050

A. Probit model of delay perception

First of all, the models described above and shown in Table 25 were run as binary probit instead of logit models to look at the impact of changes in the assumption of distribution, F (as noted in section 6.4). It is not expected that probit results would differ largely from the logit estimations reported in section 6.4. Rather, they are reported in Table 64 to investigate the sensitivity of the logit results while the predicted probabilities are shown in Figure 61. As expected, the predicted probabilities are lower for probit than logit due to logistic distribution having longer tails (Horowitz and Savin, 2001).

Table 64 Probit estimates of delay perception

	(1)	(2)	(3) - seat	(3) – no seat
Constant	-1.054^{***}	-1.078^{***}	-0.999^{***}	
	(-49.73)	(-49.90)	(-36.40)	
BnL	0.142^{***}	0.133^{***}	0.201^{***}	
	(4.32)	(3.96)	(4.76)	
CL	0.446^{***}	0.457^{***}	0.468^{***}	
	(16.37)	(16.50)	(13.11)	
CnL	0.541^{***}	0.543^{***}	0.582^{***}	
	(18.06)	(17.83)	(14.51)	
LF	-0.0286	-0.0349	-0.0121	
	(-0.99)	(-1.19)	(-0.32)	
LR	-0.00274	-0.0167	0.0248	
	(-0.11)	(-0.66)	(0.78)	
Arrival delay				
BL	0.0907^{***}	0.115^{***}	0.0745^{***}	0.0848^{**}
	(34.35)	(24.70)	(10.76)	(2.85)
BnL	0.106^{***}	0.146^{***}	0.0988^{***}	0.130^{***}
	(29.39)	(26.32)	(11.50)	(6.65)
CL	0.107^{***}	0.133^{***}	0.0805^{***}	0.139^{***}
	(35.75)	(26.76)	(9.78)	(10.06)
CnL	0.121^{***}	0.162^{***}	0.126^{***}	0.196^{***}
	(30.48)	(26.63)	(11.70)	(12.20)
LF	0.107^{***}	0.137^{***}	0.108^{***}	0.168^{***}
	(36.46)	(33.26)	(15.98)	(11.24)
LR	0.100^{***}	0.137^{***}	0.0944^{***}	0.147^{***}
	(60.22)	(51.27)	(23.48)	(13.75)
Arrival delay x SJT				
BL		-0.000206^{***}	-0.0000639	0.000352
		(-6.42)	(-1.48)	(1.06)
BnL		-0.000394^{***}	-0.000302^{***}	-0.000472^{**}
		(-10.03)	(-5.84)	(-2.69)
CL		-0.000649^{***}	-0.000164	-0.000153
		(-6.97)	(-1.24)	(-0.40)
CnL		-0.00126^{***}	-0.000997^{***}	-0.00212^{***}
		(-9.46)	(-5.82)	(-4.62)
LF		-0.000384^{***}	-0.000317^{***}	-0.000420^{*}
		(-10.80)	(-6.69)	(-2.41)
LR		-0.000302^{***}	-0.000185^{***}	-0.000381^{***}
		(-18.61)	(-8.48)	(-4.72)
Arrival delay x departure delay				
BL			0.00866^{***}	
			(13.69)	
BnL			0.00397^{***}	
			(7.61)	
CL			0.00299^{***}	
			(6.26)	
CnL			0.00137^{*}	
			(2.34)	
LF			0.00209^{***}	
			(5.83)	
LR			0.00424^{***}	
			(16.66)	
N	73050	72884	48793	
Log-likelihood	-41339.4	-40874.6	-27345.7	
Pseudo R^2	0.133	0.141	0.155	

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced.

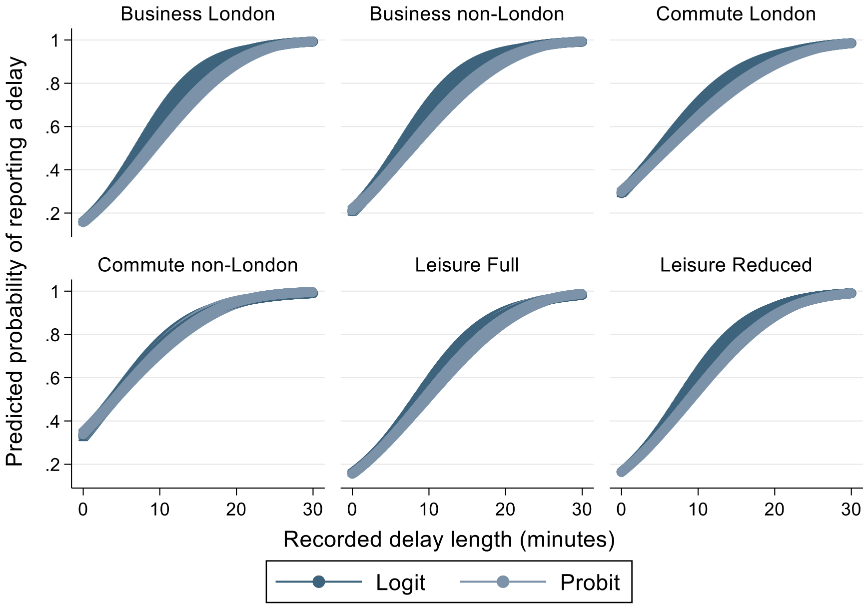


Figure 61 Predicted probabilities of delay perception for probit (model 3) under the scenario of scheduled journey time of 30 minutes, 5 minutes of delay at departure and for seated passengers

B. Interchanges and their impact on discrepancies between recorded and reported performance

As previously noted, one of the limitations of the dataset used in the analysis is the inability to have full confidence in how passenger reports of delays may be affected by delays on other journey legs or due to passengers travelling on a different service than originally planned (either previous or following services). To mitigate the possible impacts of the discrepancies resulting from the described limitations, some further analysis was conducted. The dataset was restricted to passengers using restricted (advanced) tickets to eliminate the impact of possible differences between the reported (matched) journey and the journey that a passenger planned to make. This way, there is more confidence in a passenger travelling on a service they originally planned (though this could still have been the case if some services were cancelled). Additionally, the dataset was further restricted to passengers who reported they did not change trains as part of their journey. This is to act as a sensitivity check to see whether the perception modelling results reported previously may be impacted by passengers:

a. travelling on a multi-leg journey and including a delay on the other journey legs,
b. boarding a different train than originally planned (i.e. their scheduled train departed later), this can especially happen in the case of OD pairs with frequent connections and
c. boarding a different train than originally planned (i.e. a delayed train departing earlier than the train passengers planned to travel on).

In these cases, it is impossible to deduce which applied to a passenger. Monsuur et al. (2021) looked at the impact of delays on passenger satisfaction, suggesting that passenger satisfaction is impacted negatively if the previous train was cancelled due to longer waiting times or increased crowding. However, as it is not exactly known how the travel plans differed from actual experience, there are limits as to what inferences can be drawn in this regard.

Restriction to advance tickets and only for passengers who stated that they definitely did not change trains means reducing the sample size quite substantially as the question about the interchanges was not included in all the 10 survey waves used as part of this investigation and was also characterised by a lower response rate. Additionally, the sample is only restricted to leisure travellers as this is the type of travellers more likely to choose this type of ticket. The models were, therefore, run for the restricted sample with results being presented in Table 65 and the predicted delay length thresholds (i.e. where the probability of delay perception is 0.5) are shown in Figure 62.

Table 65 Logit estimates of delay perception for a restricted sample

	(1)	(2)	(3) - seat	(3) – no seat
Constant	-2.216***	-2.270***	-2.153***	
	(-31.37)	(-31.42)	(-18.01)	
L_A				
LR	0.187***	0.250***	0.168***	0.394*
	(22.12)	(16.62)	(5.93)	(2.11)
L_A x SJT				
LR		-0.000429***	-0.000139	-0.00163
		(-5.40)	(-0.94)	(-0.86)
L_A x L_D				
LR			0.0252***	
			(6.23)	
N	3281	3279	1241	
Log-likelihood	-1559.3	-1544.6	-561.2	
Pseudo R^2	0.180	0.187	0.264	
% correct	78.94%	79.45%	81.14%	

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
LR-Leisure Reduced; ; L_A, L_D refer to arrival and departure delay; SJT – scheduled journey time

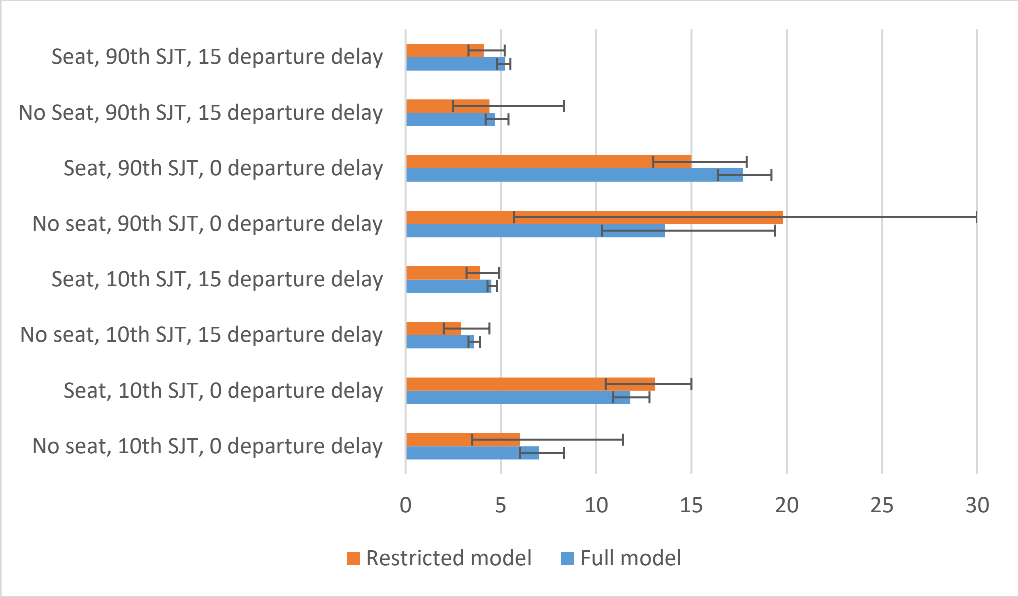


Figure 62 Delay length thresholds for P=0.5 for a restricted and full model

In terms of the predicted delay perception thresholds, one can see that the thresholds are typically smaller. However, further restricting the sample results in lower statistical significance of the estimated results and larger confidence intervals around the estimated thresholds. Moreover, as in the case of model 3, the coefficients on the interaction between arrival delay and scheduled journey time are negative but insignificant. The thresholds do not decrease significantly with increasing length of the journey as was the case with the full model.

Restricting the sample to journeys made on advance (restricted) tickets and with no interchanges aimed at improving accuracy between passenger experience, passenger reports and recorded performance. To relax this, the dataset is now only restricted to passengers not travelling on multiple journey legs, but with no restrictions related to the type of ticket. In the case of model 3, the length of arrival delay was also restricted to 25 minutes as some issues with convergence were encountered due to the small sample size in the case of longer delays. Table 66 shows the estimated coefficients for the model with the restricted sample. The main difference between the results presented as part of the sensitivity analysis and the models estimated in the main body of the thesis is the insignificance of some of the coefficients on the interaction between arrival delay and scheduled journey time in model 3. The overall sample size is still relatively large (9000 responses), but especially in the case of standing passengers, the number of responses for each of the journey purposes ranges between 30 for business London journey purpose to 300 for commute London journey purposes. This is likely affecting the statistical significance of the results from model 3 for non-interchanging passengers. Nevertheless, the estimated results are broadly in line with the main body of analysis, perhaps suggesting that the possible errors in the way that delays were recorded and satisfaction reported are not only due to the interchanges.

Table 66 Logit delay estimates of delay perception for not interchanging passengers

	(1)	(2)	(3) - seat	(3) – no seat
Constant	-1.841***	-1.893***	-1.776***	
	(-31.48)	(-31.54)	(-15.33)	
BnL	0.0753	0.0741	0.211	
	(0.78)	(0.75)	(1.13)	
CL	0.684***	0.695***	0.660***	
	(9.38)	(9.31)	(4.54)	
CnL	0.784***	0.803***	1.223***	
	(9.71)	(9.72)	(7.73)	
LF	-0.152	-0.165*	-0.0549	
	(-1.89)	(-1.99)	(-0.35)	
LR	-0.135	-0.159*	-0.140	
	(-1.91)	(-2.18)	(-1.02)	
Arrival delay				
BL	0.151***	0.196***	0.129***	0.569***
	(20.89)	(15.74)	(4.61)	(4.28)
BnL	0.202***	0.261***	0.161***	0.293**
	(16.18)	(14.12)	(3.42)	(2.68)
CL	0.191***	0.264***	0.169***	0.294***
	(23.48)	(18.90)	(4.47)	(3.84)
CnL	0.215***	0.290***	0.0567	0.271**
	(18.37)	(15.84)	(1.04)	(3.25)
LF	0.189***	0.243***	0.153***	0.391***
	(21.71)	(19.82)	(5.53)	(4.54)
LR	0.183***	0.249***	0.139***	0.366***
	(33.27)	(28.05)	(6.84)	(4.36)
Arrival delay x SJT				
BL		-0.000403***	0.0000898	-0.00447***
		(-4.66)	(0.46)	(-3.98)
BnL		-0.000610***	-0.000628	-0.000896
		(-4.81)	(-1.80)	(-1.41)
CL		-0.00171***	-0.00118	-0.00274
		(-7.01)	(-1.88)	(-1.35)
CnL		-0.00234***	-0.00159	-0.00404
		(-5.98)	(-1.69)	(-1.81)
LF		-0.000728***	-0.000654**	-0.00221
		(-6.81)	(-2.84)	(-1.73)
LR		-0.000524***	-0.000147	-0.00263*
		(-10.26)	(-1.24)	(-2.44)
Arrival delay x departure delay				
BL			0.0193***	
			(5.30)	
BnL			0.0202***	
			(4.08)	
CL			0.0321***	
			(6.29)	
CnL			0.0230***	
			(4.85)	
LF			0.0138***	
			(5.01)	
LR			0.0309***	
			(10.42)	
N	29125	29043	9075	
Log-likelihood	-15846.5	-15657.8	-4758.5	
Pseudo R^2	0.141	0.149	0.192	

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced;

C. Additional journey type categorisation

Both journey length and delay at departure previously entered the modelling as continuous variables. Additional categorisation is used to replace the continuous variables with categorical ones to aid the interpretation of the results model. This categorisation was done by Transport Focus and was based on journey type based geography, rather than reported journey purpose. These categories are:

- Airport (6.2%)
- High-speed (12.3%)
- Interurban (14.3%)
- Long commute (23.4%)
- Long distance (21.5%)
- Rural (3.6%)
- Short commute (18.7%)

The aforementioned categorisation may serve as an additional categorisation of journey type used to control for the impacts of possible divergence between the planned and actual journey. For example, the frequency of services in urban areas is typically higher than in rural areas, meaning that in such cases passengers may be more likely to travel on a different than planned service. It is believed that the inclusion of the outlined categories may provide some additional insights and help control for any likely discrepancies.

The geographical journey types are likely characterised by some inherent differences that relate to the type of service – both in terms of journey lengths, headways and journey quality. Short commute, airport and long commute journeys are all characterised by an average journey time of below 50 minutes. Rural and interurban services are characterised by an average journey time close to the sample average while high-speed and long-distance journeys have average journey times of around 100 minutes. Therefore, the initial version of the model (i.e. model 1) was estimated with the arrival delay coefficient being estimated separately for each of the journey purpose and journey type combinations. The estimated coefficients are shown in Table 67 with the graphical representation shown in Figure 63. This was not done for models 2 and 3 as it is believed that the additional categorisation would reduce sample sizes – 7 journey type categories, 6 journey purposes and 2 journey quality categories would ultimately lead to 84 estimated coefficients with each of the groups likely having a reduced sample size, affecting the significance of the results. Moreover, the journey type categories, as mentioned before, are likely to be correlated with some of the explanatory variables used in models 2 and 3 as discussed above in the case of scheduled journey time.

Table 67 Logit estimates of delay perception for a model with typology

Constant	-1.798***
	(-46.42)
BL	0
	(.)
BnL	0.240***
	(4.05)
CL	0.759***
	(15.57)
CnL	0.845***
	(15.60)
LF	-0.0225
	(-0.43)
LR	0.00725
	(0.16)
L_A – Airport	
BL	0.104***
	(9.55)
BnL	0.245**
	(2.62)
CL	0.101***
	(8.48)
CnL	0.332***
	(3.69)
LF	0.114***
	(10.44)
LR	0.145***
	(8.97)
L_A – High-speed	
BL	0.171***
	(25.31)
BnL	0.237***
	(12.18)
CL	0.158***
	(11.92)
CnL	0.179***
	(10.46)
LF	0.167***
	(13.43)
LR	0.169***
	(33.47)
L_A – Interurban	
BL	0.185***
	(13.71)
BnL	0.162***
	(16.76)
CL	0.212***
	(12.12)
CnL	0.212***
	(16.74)
LF	0.197***
	(20.28)
LR	0.190***
	(32.12)
L_A – Long commute	
BL	0.224***

	(23.86)
BnL	0.310***
	(12.86)
CL	0.200***
	(28.61)
CnL	0.243***
	(17.47)
LF	0.230***
	(21.88)
LR	0.211***
	(33.77)
L_A – Long distance	
BL	0.117***
	(19.52)
BnL	0.184***
	(22.70)
CL	0.208**
	(3.29)
CnL	0.218***
	(18.57)
LF	0.185***
	(23.03)
LR	0.155***
	(42.41)
L_A – Rural	
BL	-
	-
BnL	0.179***
	(9.54)
CL	-
	-
CnL	0.179***
	(8.54)
LF	0.166***
	(11.43)
LR	0.188***
	(16.78)
L_A – Short commute	
BL	0.257***
	(9.16)
BnL	0.204***
	(10.82)
CL	0.199***
	(24.52)
CnL	0.276***
	(20.92)
LF	0.214***
	(18.18)
LR	0.226***
	(19.52)
N	73050
Log-likelihood	-40987.3
Pseudo R^2	0.140
% correct	72.07%

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced; L_A refers to arrival and departure delay; - no observations

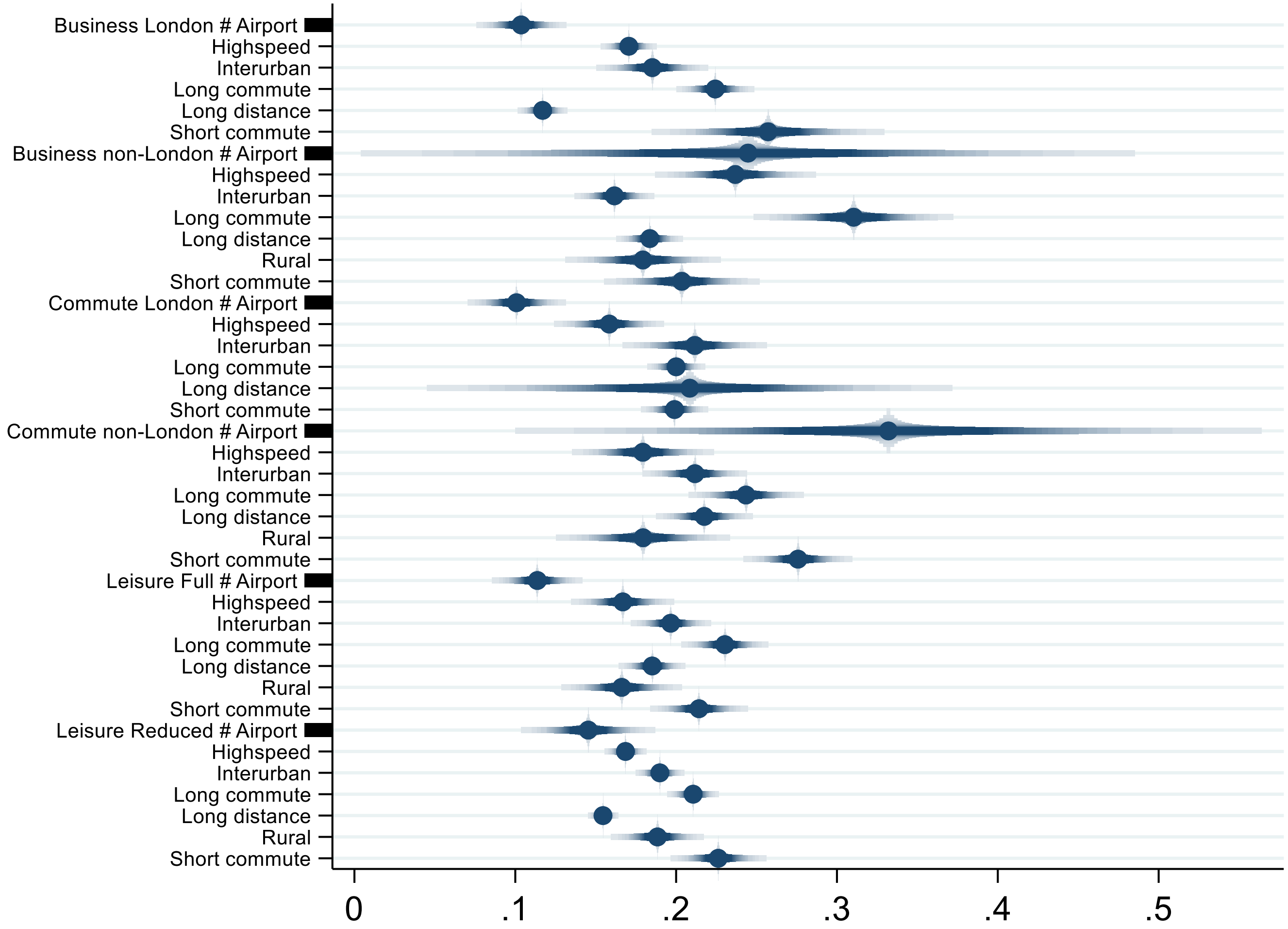


Figure 63 Coefficient plot for the logit model of delay perception by journey purpose and typology

Looking at the sizes of arrival delay length coefficients for different journey types within the same journey purpose, it can be investigated whether there are any significant differences in the impacts of delay length on delay perception that result from the specific journey type characteristics.

Business

For business London travellers, the coefficients generally become smaller for journey types with lengthier average journey times. This is especially evident in the case of long-distance journeys where the predicted delay length perception threshold is 17 minutes compared to 7-8 for short and long commute. For non-London business travellers, high-speed and long commute are characterised by larger coefficients than other journey types with the predicted delay length perception thresholds being at 5-6.5 minutes compared to 8-9 for the other journey types. It is not immediately clear why such differences between London and non-London travellers would be present. However, it is expected that differences in headways, service quality or journey times may play an important role in determining these differences.

Commute

For commuters, the differences between journey types are not significant. In the case of London commuters, a slightly lower coefficient for high-speed may indicate on the impact of journey quality on delay perception. In the case of non-London commuters, short and long commute are characterised by larger coefficients (the difference is only statistically significant in the case of short commutes). The delay perception thresholds are between 3-4 minutes for the short and long commute and 4-5 for the other journey types. It is, however, worth mentioning that for non-London commute, the average journey times for all the journey types are very similar – between 21.4 and 29.7 minutes. This is unlike for the other journey types where the average journey times typically differ largely by journey type.

Leisure

For both leisure travellers on full and reduced ticket types, the pattern in the coefficient sizes is very similar. High-speed, long-distance, rural and interurban journey types are characterised by smaller coefficients whereas these journeys are also on average longer. At the same time, the coefficients for short and long commute are typically larger. The predicted delay length perception thresholds range between 9-12 minutes for the journey types characterised by lengthier journeys and around 8 minutes for the shorter journeys. This highlights the importance of journey length in determining the delay perception for leisure travellers with this result being generally consistent with what was reported in the main body of the thesis.

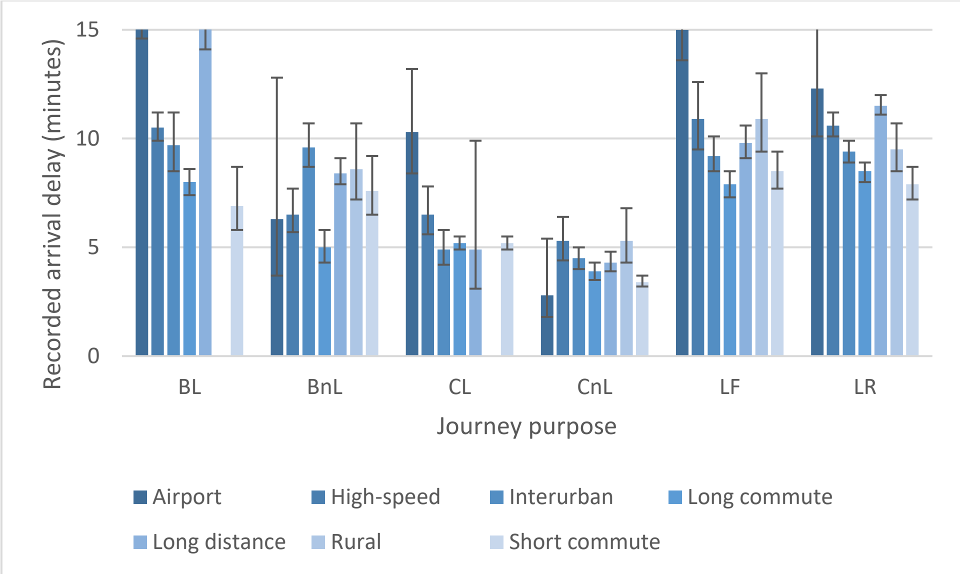


Figure 64 Delay length thresholds at p=0.5 by journey purpose and typology

While the results of this analysis are in some cases difficult to explain or the differences are not statistically significant, they provide additional insights into the determinants of delay perception. It needs to be noted that journey types may have some characteristics that are not measured or quantifiable, i.e. other than journey time that may have an impact on how delays are perceived. These can generally relate to differences in journey quality, headways or in-station and on-board announcements.

D. Summary

The purpose of this annex was to provide sensitivity analysis around the previously estimated models of delay perception. This was done due to the novelty of the analysis as previous research into how passengers perceive delays is very limited. Moreover, as the NRPS dataset was suggested to have some limitations related to possible differences between passengers' travel plans, actual journeys and how these are recorded in the questionnaire, an attempt was made to restrict the dataset to responses where the scope for error is limited (i.e. travellers with restricted tickets and no interchanges). Furthermore, additional journey type categorisation was introduced to better understand how different types of passengers may perceive delays as some differences may be expected due to different types of journeys being characterised by different journey lengths, frequencies, comfort of journey or delay sensitivities. Most of the results are in line with the main body of analysis, though, due to lower sample sizes, the significance of results is typically smaller.

Annex II

Interaction models

This annex presents supplementary analysis comparing results from the preferred models of delay perception and satisfaction to their re-esimated versions that include all levels of interacted variables.

A. Binary logit model of delay perception from Table 25

Average marginal effects are reported in Table 68. The predicted probabilities are reported in Figure 65 and Figure 66 as well as Table 69.

Table 68 Average marginal effects for model 2 and 2a from Table 25

	(2)	(2a)		(2)	(2a)
Arrival delay			**SJT**		
BL	0.0303***	0.0303***	BL	-0.00034***	-0.00039***
BnL	0.0404***	0.0399***	BnL	-0.00066***	-0.00097***
CL	0.0343***	0.0344***	CL	-0.00120***	-0.00121***
CnL	0.0329***	0.0361***	CnL	-0.00217***	-0.00255***
LF	0.0338***	0.0339***	LF	-0.00064***	-0.00094***
LR	0.0360***	0.0353***	LR	-0.00049***	-0.00092***

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced

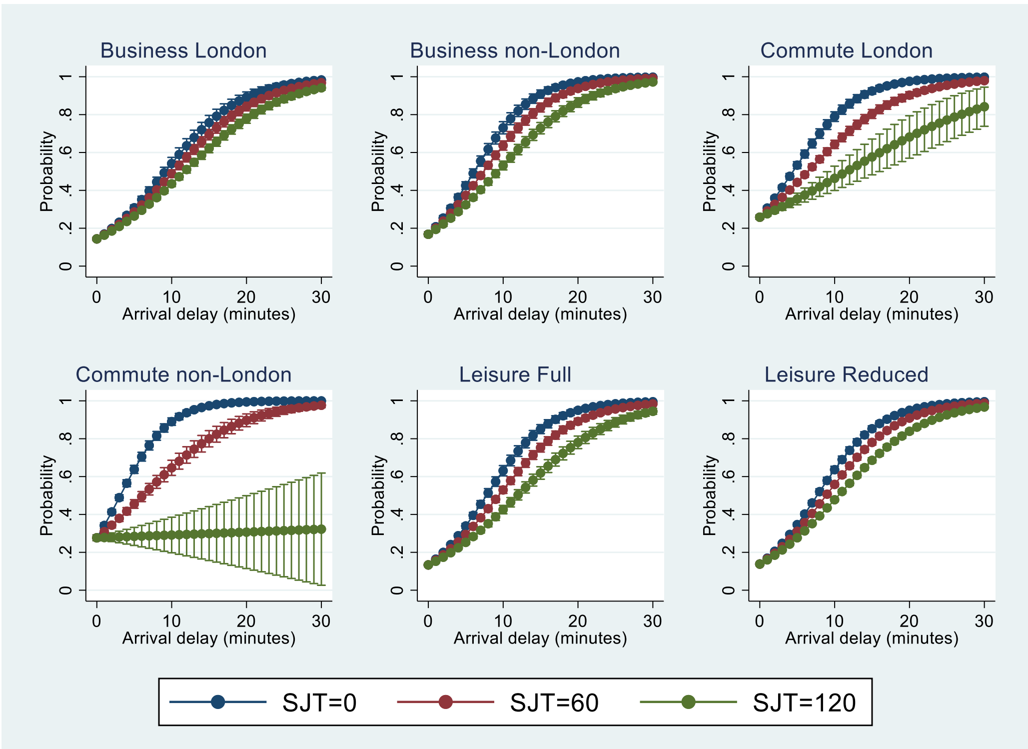


Figure 65 Probability of delay perception for increasing delay lengths, scheduled journey lengths and different journey purposes using model 2

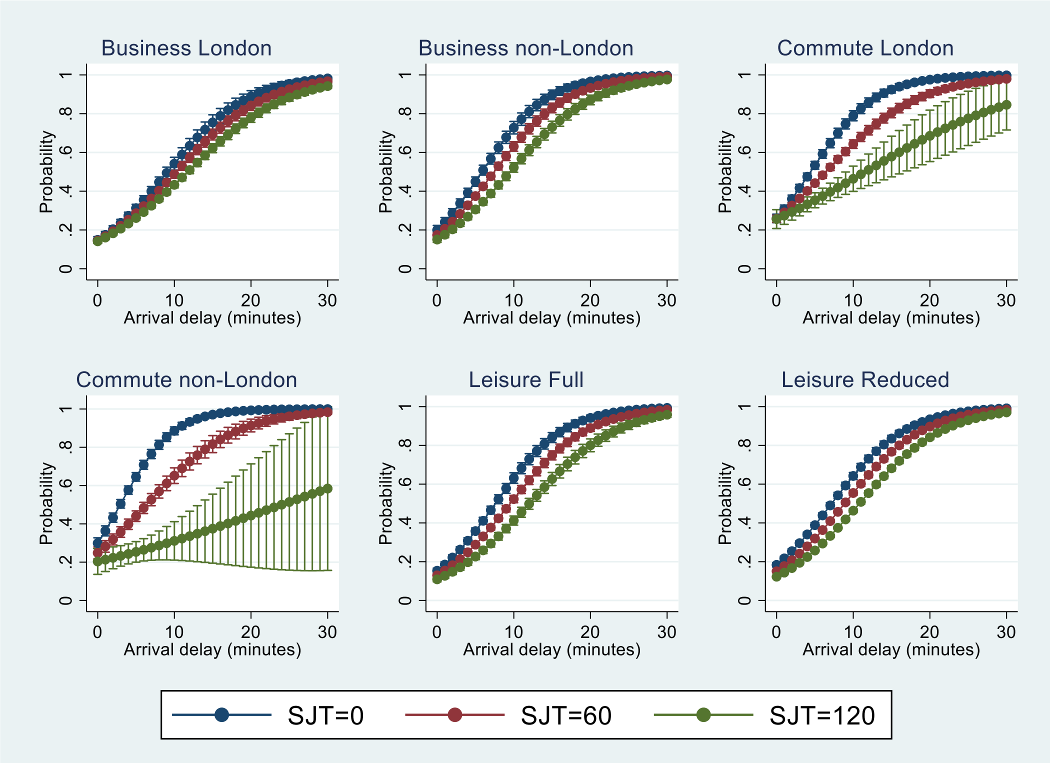


Figure 66 Probability of delay perception for increasing delay lengths, scheduled journey lengths and different journey purposes using model 2a with all levels of interacted variables

Table 69 Estimated probabilities of delay perception for different levels of scheduled journey time and journey purposes using model 2 and model 2a

		Model 2		Model 2a	
	SJT	Prob	95% CI	Prob	95% CI
Business London	30	0.31	0.30, 0.32	0.31	0.30, 0.33
Business London	90	0.29	0.28, 0.30	0.29	0.28, 0.30
Business non-London	30	0.39	0.38, 0.40	0.40	0.39, 0.42
Business non-London	90	0.35	0.34, 0.36	0.35	0.34, 0.36
Commute London	30	0.47	0.46, 0.48	0.47	0.46, 0.48
Commute London	90	0.40	0.38, 0.42	0.40	0.37, 0.42
Commute non-London	30	0.51	0.50, 0.52	0.51	0.50, 0.52
Commute non-London	90	0.37	0.33, 0.41	0.35	0.31, 0.39
Leisure Full	30	0.33	0.32, 0.34	0.33	0.32, 0.34
Leisure Full	90	0.29	0.28, 0.30	0.28	0.27, 0.29
Leisure Reduced	30	0.34	0.33, 0.34	0.36	0.35, 0.37
Leisure Reduced	90	0.31	0.30, 0.31	0.30	0.30, 0.31

Model 3 extended model 2 by addition of more explanatory variables and interactions (Table 26). Average marginal effects are reported in Table 70 for both models with the estimated probabilities shown in Figure 67 and Figure 68 for increasing levels of delay at arrival and departure.

Table 70 Average marginal effects for models reported in Table 26

	(3)	(3a)
Arrival delay		
BL	0.0268***	0.0182***
BnL	0.0271***	0.0177***
CL	0.0304***	0.0211***
CnL	0.0237***	0.0177***
LF	0.0263***	0.0153***
LR	0.0270***	0.0183***
SJT		
BL	-0.0000691	0.0000402
BnL	-0.000439***	-0.000516***
CL	-0.000161	0.00116***
CnL	-0.00178***	-0.000559
LF	-0.000445***	-0.000367***
LR	-0.000269***	-0.000343***
Departure delay		
BL	0.0167***	0.0595***
BnL	0.0137***	0.0607***
CL	0.0109***	0.0704***
CnL	0.00684***	0.0671***
LF	0.00928***	0.0531***
LR	0.0126***	0.0566***
Seat=1		
BL	-0.110***	-0.0822**
BnL	-0.0889**	-0.0464
CL	-0.138***	-0.172***
CnL	-0.121***	-0.127**
LF	-0.143***	-0.0985***
LR	-0.155***	-0.0982***

Legend: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$;
BL/BnL – Business London/non-London, CL/CnL – Commute London/non-London, LF/LR-Leisure Full/Reduced

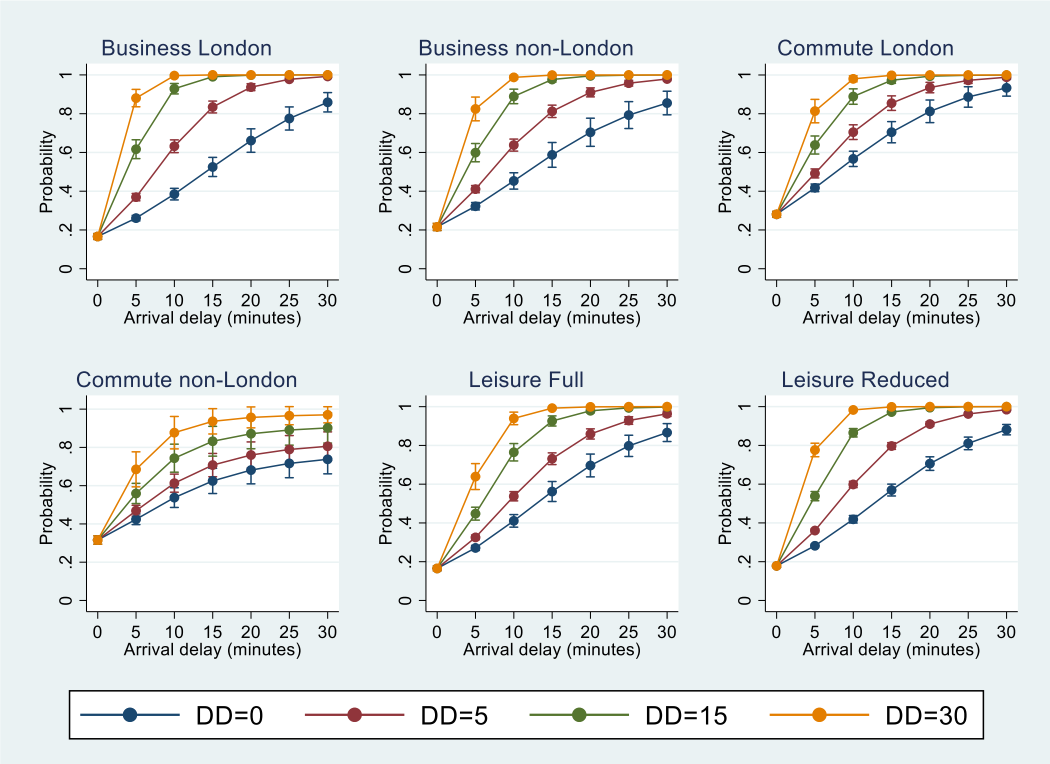


Figure 67 Probability of delay perception for increasing delay lengths and different journey purposes using model 3 in Table 26 (DD refers to departure delay)

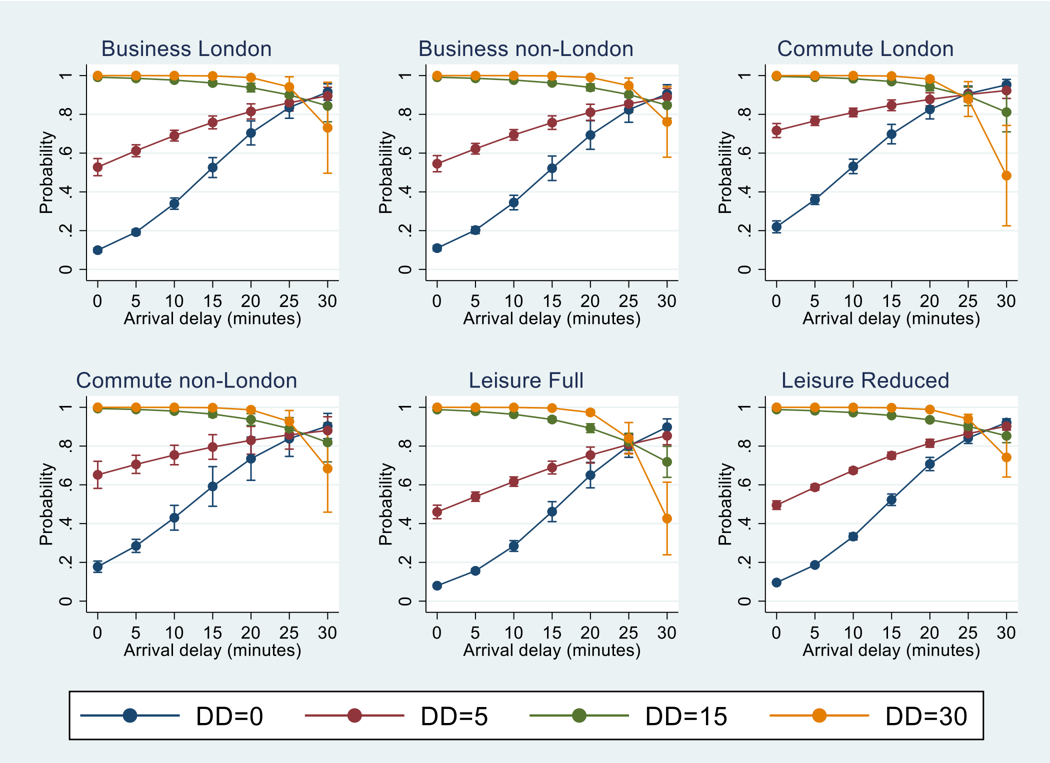


Figure 68 Probability of delay perception for increasing delay lengths and different journey purposes using model 3a in Table 26 (DD refers to departure delay)

B. Binary logit model of punctuality satisfaction with version 2 of the satisfaction classification from Table 36

Average marginal effects are reported in Table 71. The estimated probabilities are then plotted in Figure 69 and Figure 70.

Table 71 Average marginal effects for models from Table 36

	(V2)	(V2a)
Arrival delay		
BL	-0.0147^{***}	-0.0102^{***}
BnL	-0.0199^{***}	-0.0112^{***}
CL	-0.0257^{***}	-0.0186^{***}
CnL	-0.0317^{***}	-0.0185^{***}
LF	-0.0160^{***}	-0.00850^{***}
LR	-0.0131^{***}	-0.00780^{***}
SJT		
BL	0.000227^{***}	0.000334^{***}
BnL	0.000282^{***}	0.000483^{***}
CL	0.000536^{***}	-0.000243
CnL	0.000756^{***}	-0.000159
LF	0.000360^{***}	0.000400^{***}
LR	0.000186^{***}	0.000348^{***}
Departure delay		
BL	-0.00343^{***}	-0.0237^{***}
BnL	-0.00220^{***}	-0.0289^{***}
CL	-0.00107^{*}	-0.0442^{***}
CnL	0.000889^{*}	-0.0465^{***}
LF	-0.000999^{***}	-0.0215^{***}
LR	-0.00236^{***}	-0.0214^{***}
Seat=1		
BL	0.0367^{***}	0.119^{***}
BnL	0.0424^{***}	0.0914^{***}
CL	0.0987^{***}	0.236^{***}
CnL	0.0459	0.233^{***}
LF	0.0363^{***}	0.0892^{***}
LR	0.0391^{***}	0.0996^{***}

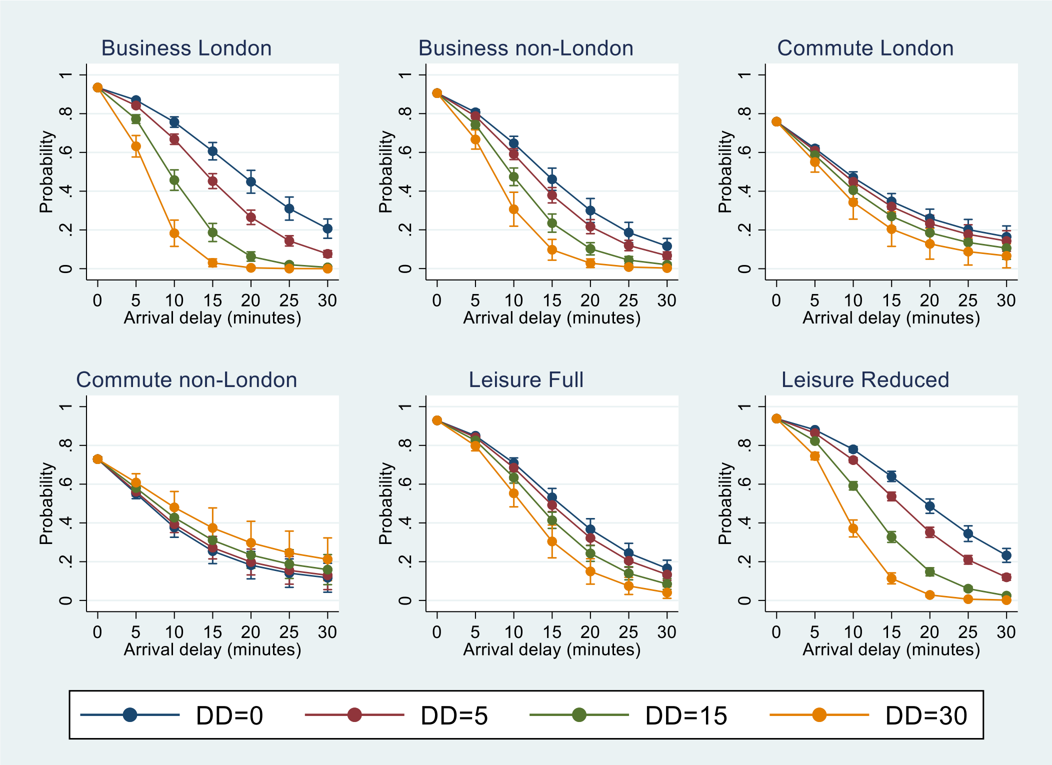


Figure 69 Estimated probability of punctuality satisfaction for increasing lengths of departure (DD) and arrival delays, using model V2 in Table 36

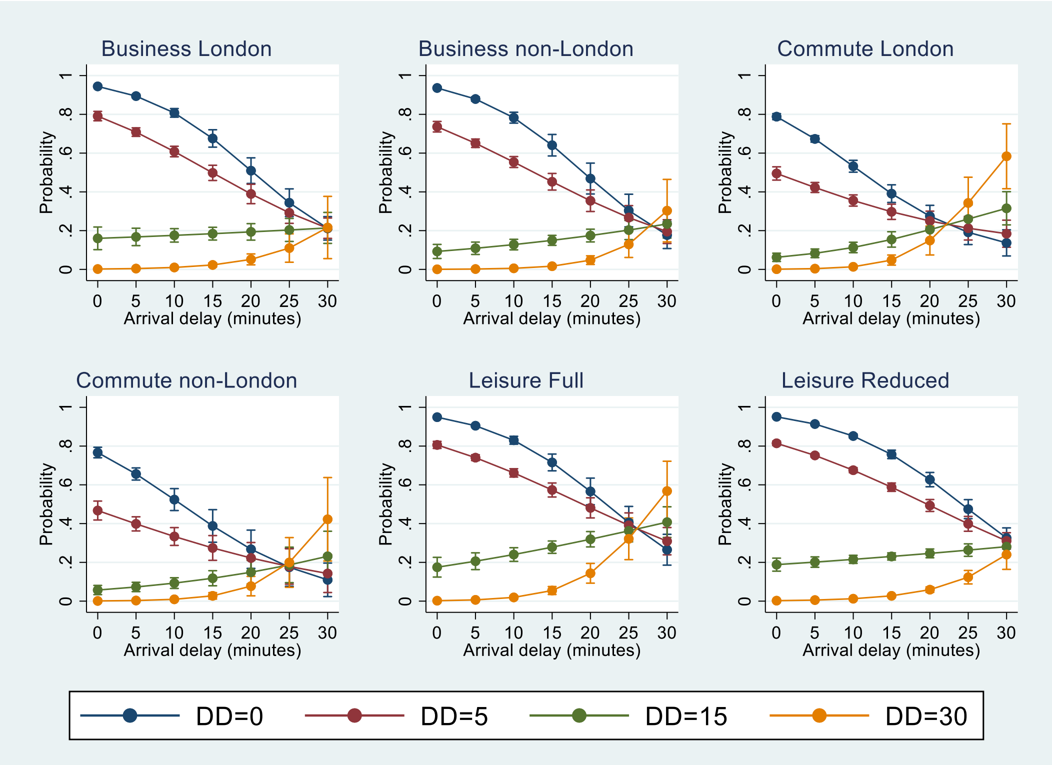


Figure 70 Estimated probability of punctuality satisfaction for increasing lengths of departure (DD) and arrival delays, using model V2a in Table 36

C. Ordered logit model of punctuality satisfaction from Table 41

It is generally more difficult to depict the results of such a model as with 5 choice categories and multiple interactions. Three sets of plots are shown below for the original and the re-estimated model with all the levels of interacted variables for increasing levels of delay at arrival and delay at departure of respectively 0, 15 and 30 minutes in Figure 71, Figure 72 and Figure 73 for the original version of the model, and Figure 74, Figure 75 and Figure 76 for the model estimated with all levels of the interacted variables.

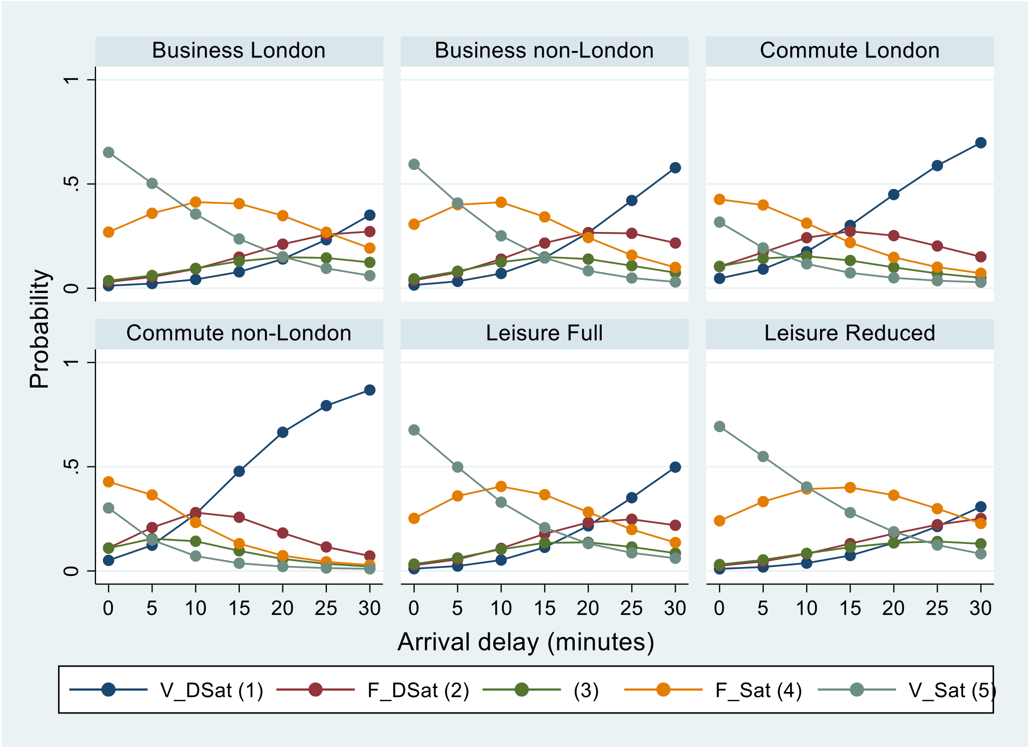


Figure 71 Probability of punctuality satisfaction based on the ordered logit model for increasing lengths of arrival (based on the model Punc_Sat from Table 41)

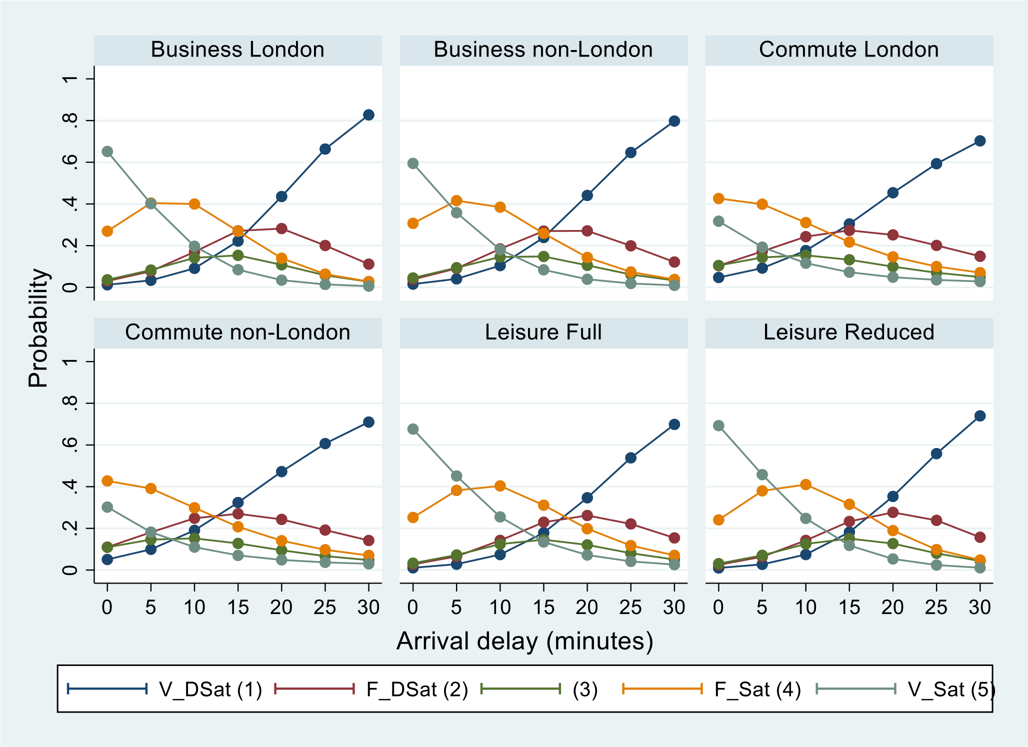


Figure 72 Probability of punctuality satisfaction based on the ordered logit model for increasing lengths of arrival delay and departure delay of 15 minutes (based on the model Punc_Sat from Table 41)

Figure 73 Probability of punctuality satisfaction based on the ordered logit model for increasing lengths of arrival delay and departure delay of 30 minutes (based on the model Punc_Sat from Table 41)

Figure 74 Probability of punctuality satisfaction based on the ordered logit model for increasing lengths of arrival delay and departure delay of 0 minutes (based on the model Punc_Sat_1 from Table 41)

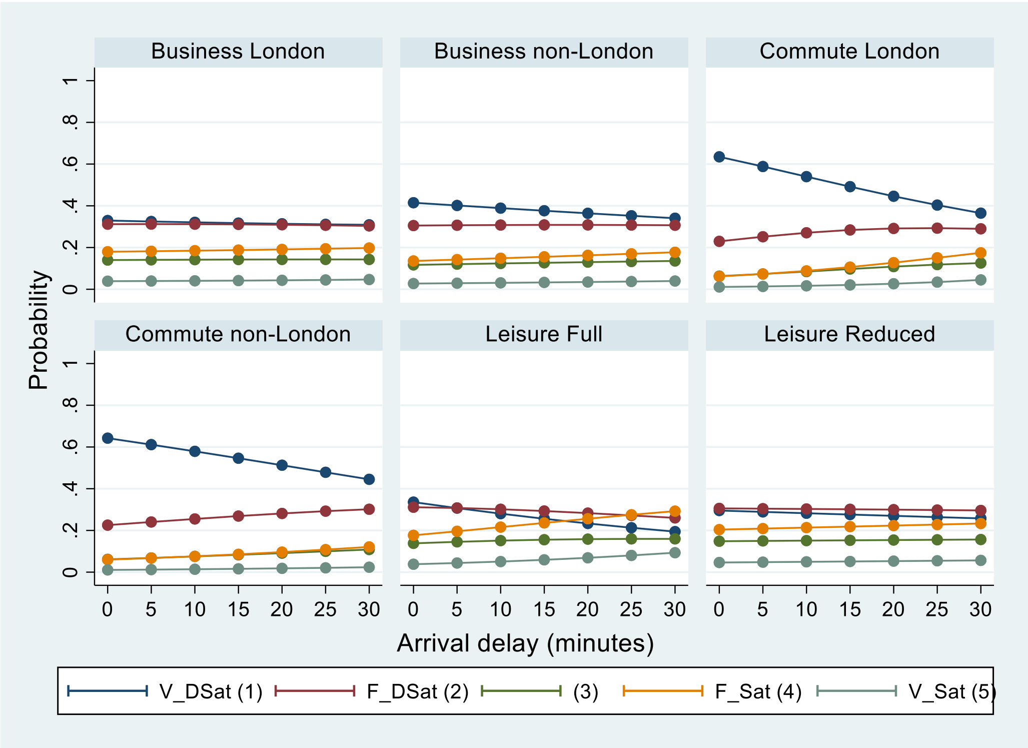


Figure 75 Probability of punctuality satisfaction based on the ordered logit model for increasing lengths of arrival delay and departure delay of 15 minutes (based on the model Punc_Sat_1 from Table 41)

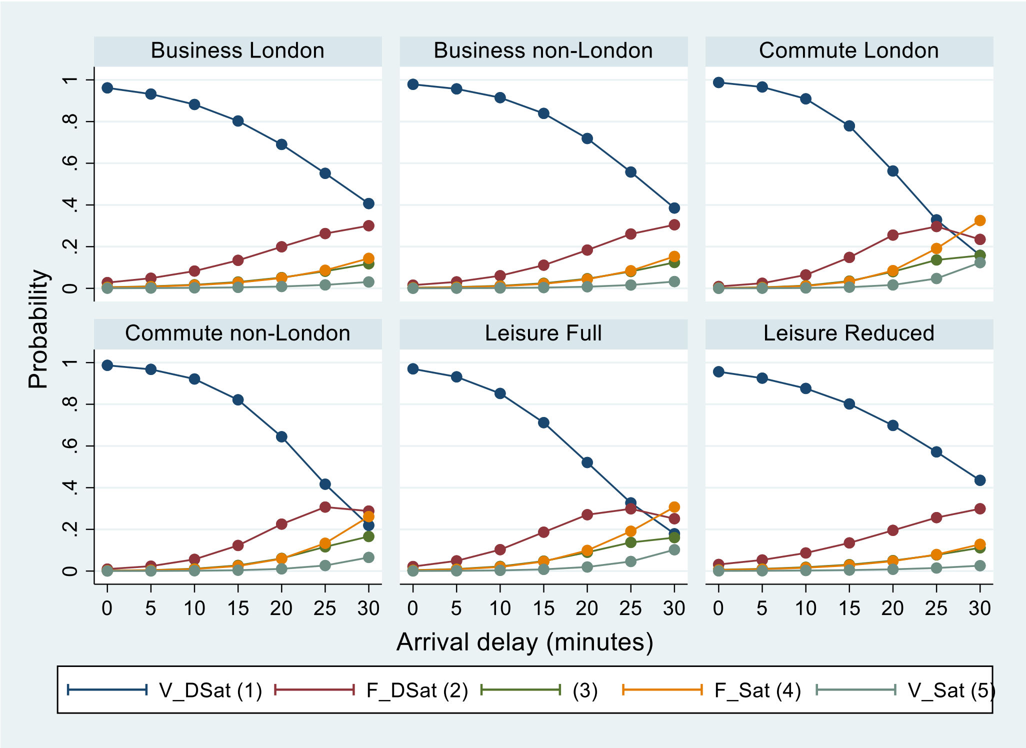


Figure 76 Probability of punctuality satisfaction based on the ordered logit model for increasing lengths of arrival delay and departure delay of 30 minutes (based on the model Punc_Sat_1 from Table 41)

D. Fractional logit OD model of punctuality satisfaction from Table 45

The average marginal effects are reported in Table 72. Figure 77 compares the estimated proportions of satisfied passengers using the two versions of the estimated models for different levels of APL and journey purposes. Figure 78, Figure 79 and Figure 80 then show the estimated proportions of satisfied passengers for different journey purposes, levels of APL and scheduled journey times using V2 of the binary representation of the satisfaction variable.

Table 72 Average marginal effects for models from Table 45

	V2	V2_AL
Sat_V1		
APL	-0.0389	-0.0392
SJT	0.0002	0.0001
PSeat	0.2516	0.3714
Sat_V2		
APL	-0.0398	-0.0395
SJT	0.0006	0.0005
PSeat	0.1285	0.2537
Sat_V3		
APL	-0.0313	-0.0311
SJT	0.0005	0.0005
PSeat	0.0901	0.1668

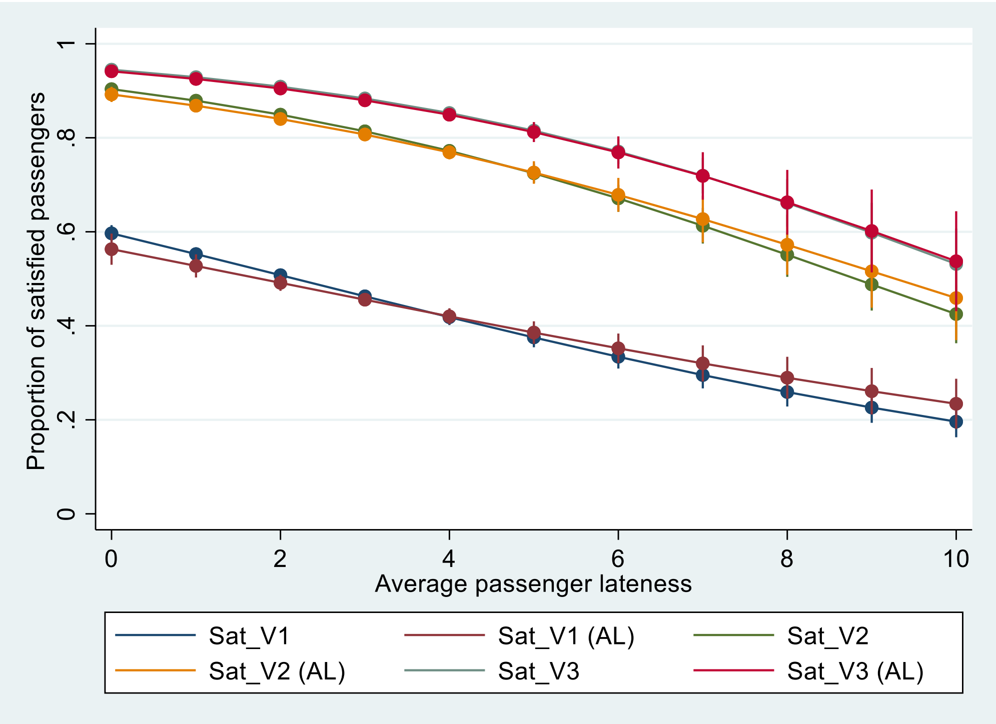


Figure 77 Levels of 'delay satisfaction' for increasing average delay , V1: (5) vs (1-4); V2: (4-5) vs (1-3); V3: (3-5) vs (1-2) based on the model from Table 45 (AL refers to models with all levels of the interacted variables)

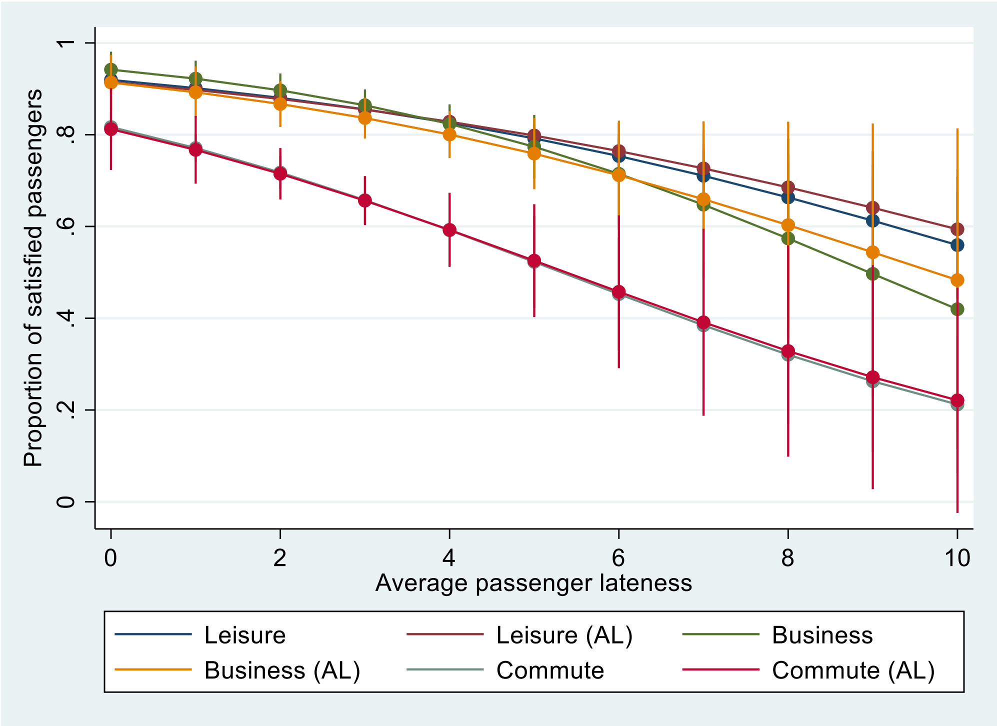


Figure 78 Proportion of satisfied passengers under Version 2 of delay satisfaction at the average values of control variables for the original model and model with all levels of interacted variables (AL) from Table 45

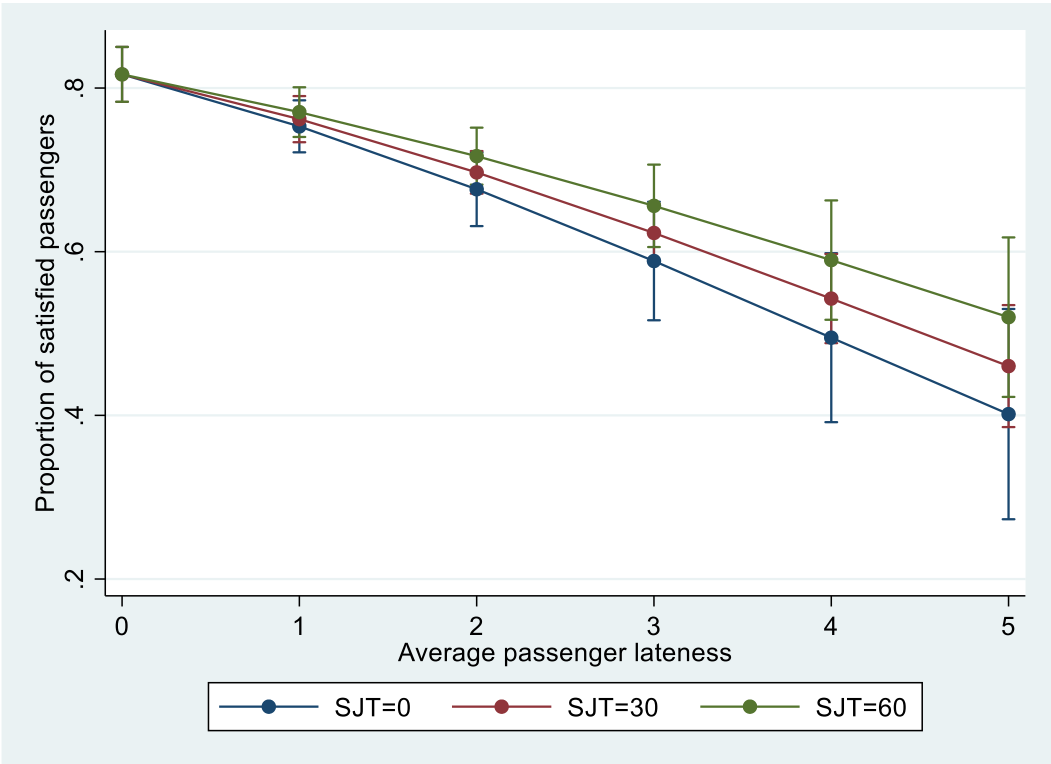


Figure 79 Proportion of satisfied passengers using V2 of the satisfaction variable and results of the model V2 from Table 45

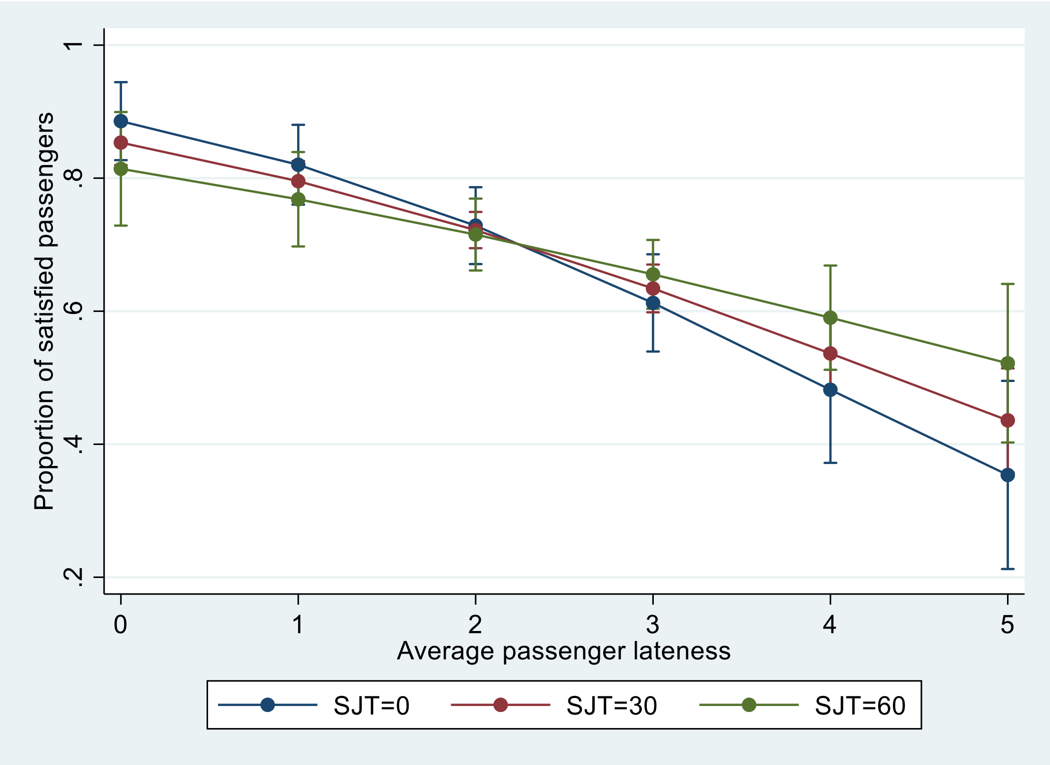


Figure 80 Proportion of satisfied passengers using V2 of the satisfaction variable and results of the model V2_AL from Table 45

E. Binary models of delay perception and punctuality dissatisfaction from Table 47

Average marginal effects are presented in Table 73. The estimated probabilities are compared in Table 74 and Table 75 and are plotted in Figure 81 and Figure 82.

Table 73 Average marginal effects for models from Table 47

	Perc	Perc_AL	DSat	DSat_AL
Arrival delay				
Business	0.0329***	0.0321***	0.0201***	0.0192***
Commute	0.0405***	0.0417***	0.0268***	0.0282***
Leisure	0.0335***	0.0323***	0.0178***	0.0170***
Seat				
Business	-0.0668***	-0.123***	-0.0968***	-0.192***
Commute	-0.0928***	-0.143***	-0.113***	-0.184***
Leisure	-0.0914***	-0.174***	-0.0925***	-0.191***

Table 74 Estimated probabilities of delay perception using models from Table 47

Journey purpose	Seat	Arrival delay	Perc	95%CI	Perc_AL	95% CI
Business	0	0	0.18	0.17,0.19	0.28	0.23,0.33
Business	0	5	0.41	0.38,0.45	0.46	0.42,0.51
Business	0	10	0.70	0.63,0.76	0.66	0.60,0.72
Business	1	0	0.18	0.17,0.19	0.17	0.16,0.18
Business	1	5	0.33	0.32,0.34	0.32	0.31,0.33
Business	1	10	0.52	0.51,0.54	0.52	0.51,0.54
Commute	0	0	0.29	0.28,0.30	0.41	0.38,0.43
Commute	0	5	0.64	0.62,0.66	0.65	0.63,0.67
Commute	0	10	0.89	0.87,0.91	0.84	0.81,0.86
Commute	1	0	0.29	0.28,0.30	0.27	0.25,0.28
Commute	1	5	0.50	0.49,0.51	0.49	0.48,0.50
Commute	1	10	0.71	0.69,0.72	0.72	0.70,0.74
Leisure	0	0	0.16	0.15,0.16	0.31	0.27,0.34
Leisure	0	5	0.42	0.40,0.45	0.50	0.47,0.53
Leisure	0	10	0.75	0.71,0.79	0.69	0.65,0.73
Leisure	1	0	0.16	0.15,0.16	0.15	0.14,0.16
Leisure	1	5	0.30	0.30,0.31	0.30	0.29,0.30
Leisure	1	10	0.51	0.50,0.52	0.51	0.50,0.52

Table 75 Estimated probabilities of delay dissatisfaction using model from Table 47

Journey purpose	Seat	Arrival delay	DSat	95% CI	DSat_AL	95% CI
Business	0	0	0.11	0.11,0.12	0.26	0.22,0.31
Business	0	5	0.29	0.26,0.31	0.39	0.35,0.43
Business	0	10	0.55	0.49,0.62	0.53	0.48,0.59
Business	1	0	0.11	0.11,0.12	0.11	0.10,0.11
Business	1	5	0.19	0.18,0.20	0.18	0.18,0.19
Business	1	10	0.30	0.29,0.31	0.30	0.29,0.31
Commute	0	0	0.34	0.33,0.35	0.49	0.46,0.52
Commute	0	5	0.62	0.60,0.64	0.65	0.63,0.67
Commute	0	10	0.83	0.81,0.86	0.78	0.75,0.80
Commute	1	0	0.34	0.33,0.35	0.30	0.29,0.32
Commute	1	5	0.46	0.45,0.47	0.45	0.44,0.46
Commute	1	10	0.59	0.58,0.61	0.60	0.59,0.62
Leisure	0	0	0.09	0.09,0.10	0.24	0.21,0.27
Leisure	0	5	0.24	0.22,0.26	0.36	0.33,0.38
Leisure	0	10	0.49	0.45,0.54	0.49	0.46,0.53
Leisure	1	0	0.09	0.09,0.10	0.09	0.08,0.09
Leisure	1	5	0.16	0.15,0.16	0.15	0.15,0.16
Leisure	1	10	0.26	0.25,0.26	0.25	0.24,0.26

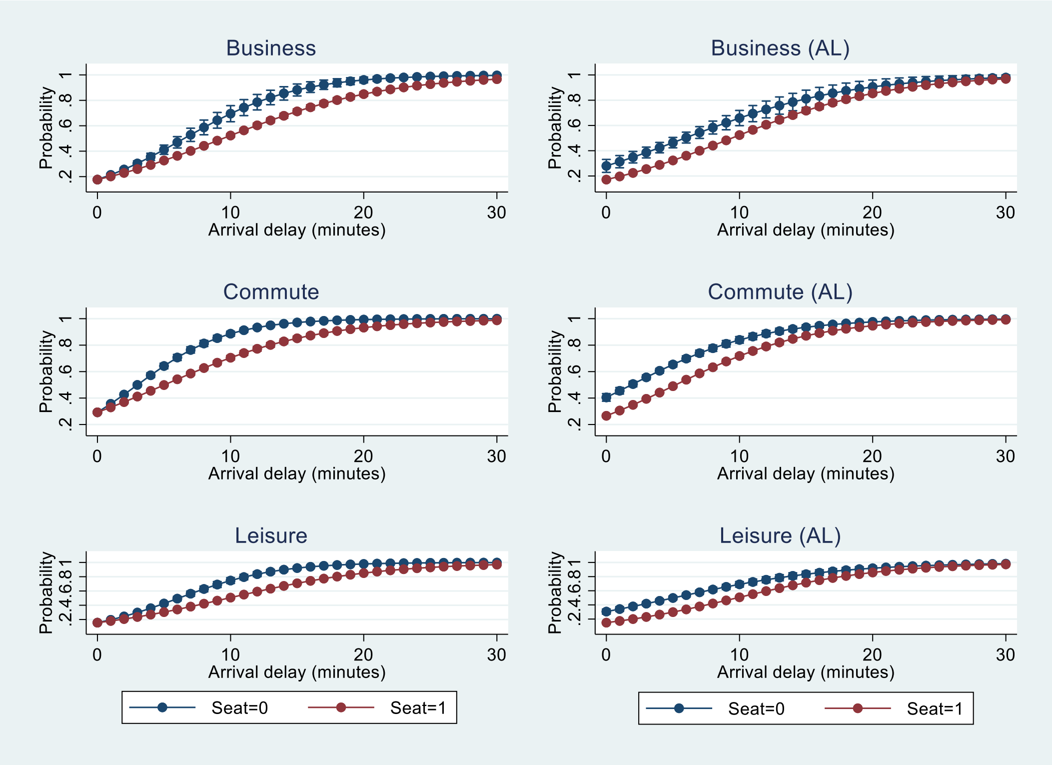


Figure 81 Estimated probabilities of delay perception using models Perc and Perc_AL from Table 47

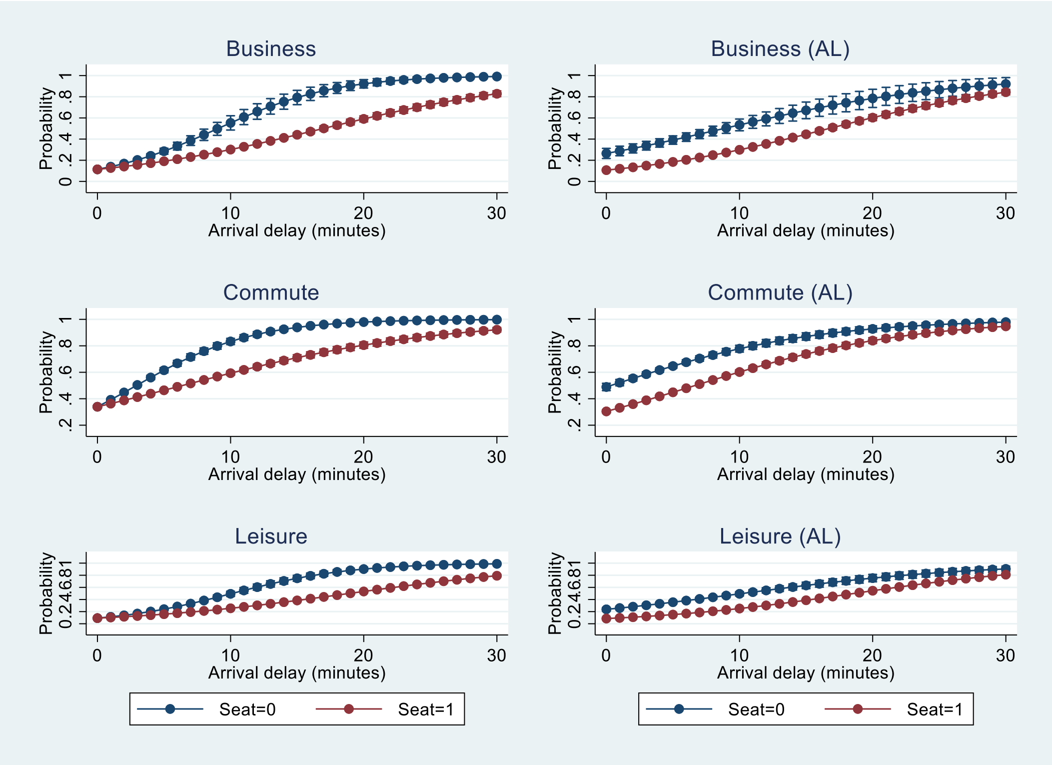


Figure 82 Estimated probabilities of delay dissatisfaction using models DSat and DSat_AL from Table 47

References

Abenoza, R.F., Cats, O. and Susilo, Y.O. 2019. Determinants of traveler satisfaction: Evidence for non-linear and asymmetric effects. *Transportation Research Part F: Traffic Psychology and Behaviour*. **66**, pp.339–356.

Abenoza, R.F., Cats, O. and Susilo, Y.O. 2017. Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transportation Research Part A: Policy and Practice*. **95**, pp.64–84.

Anthoff, D., Tol, R., Yohe, G., CT and USA 2009. Discounting for climate change. *Economics: The Open-Access, Open-Assessment E-Journal*. **3**(24).

APF 2015. *Fahrgastrechte-Statistik 2015*. Vienna.

APF 2018. *Fahrgastrechte-Statistik 2018*. Vienna.

Armstrong, J. and Preston, J. 2017. Capacity utilisation and performance at railway stations. *Journal of Rail Transport Planning & Management*. **7**(3), pp.187–205.

ATOC 2004. *Passenger Demand Forecasting Handbook (PDFH)*.

Bachmann, T.M. 2020. Considering environmental costs of greenhouse gas emissions for setting a CO2 tax: A review. *Science of The Total Environment*. **720**, p.137524.

Baetschmann, G., Staub, K.E. and Winkelmann, R. 2015. Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society. Series A: Statistics in Society*. **178**(3), pp.685–703.

Bailey, W.C., Martin, J.D. and Gray, L.N. 1974. Crime and deterrence: A correlation analysis. *Journal of Research in Crime and Delinquency*. **11**(2), pp.124–143.

Balcombe, R.J., York, I.O. and Webster, D.C. 2003. Factors influencing trip mode choice. *TRL Report*. **568**.

Baltagi, B. H. 2021. *Econometric Analysis of Panel Data*. Springer International Publishing.

Bates, J., Polak, J., Jones, P. and Cook, A. 2001. The valuation of reliability for personal travel. *Transportation Research Part E*. **37**(2–3), pp.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**(4), pp.387–408.

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M.O., Chintakayala, P.K., Choudhury, C., Daly, A., Dekker, T., Drivyla, E., Fowkes, T., Hess, S., Heywood, C., Johnson, D., Laird, J., Mackie, P., Parkin, J., Sanders, S., Sheldon, R., Wardman, M. and Worsley, T. 2019. New appraisal values of travel time saving and reliability in Great Britain. *Transportation*. **46**(3), pp.583–621.

Batley, R., Dargay, J. and Wardman, M. 2011. The impact of lateness and reliability on passenger rail demand. *Transportation Research Part E: Logistics and Transportation Review*. **47**(1), pp.61–72.

Batley, R. and Ibáñez, J.N. 2012. Randomness in preference orderings, outcomes and attribute tastes: An application to journey time risk. *Journal of Choice Modelling*. **5**(3), pp.157–175.

Becker, G.S. 1965. A theory of the allocation of time. *The Economic Journal*. **75**(299), p.493.

Bergström, A. and Krüger, N. 2013. Modeling passenger train delay distributions : evidence and implications. *Working Papers in Transport Economics 2013*. **3**.

Bernoulli, D. 2011. Exposition of a new theory on the measurement of risk. *The Kelly Capital Growth Investment Criterion: Theory And Practice*. **22**(1), pp.11–24.

Bhadra, D. 2009. You (expect to) get what you pay for: A system approach to delay, fare, and complaints. *Transportation Research Part A: Policy and Practice*. **43**(9), pp.829–843.

Bhat, C.R. and Sardesai, R. 2006. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological*. **40**(9), pp.709–730.

Boes, S. and Winkelmann, R. 2006. Ordered response models. *Allgemeines Statistisches Archiv*. **90**, pp.167–181.

Bonera, M. and Martinelli, V. 2023. Covid-19 and Public Transport: two years later. Investigating the transport demand trend in the City of Brescia. *Transportation Research Procedia*. **69**, pp.376–383.

Börjesson, M. and Eliasson, J. 2011. On the use of ‘ average delay’ as a measure of train reliability. *Transportation Research Part A: Policy and Practice*. **45**(3), pp.171–184.

Börjesson, M. and Rubensson, I. 2019. Satisfaction with crowding and other attributes in public transport. *Transport Policy*. **79**, pp.213–222.

Boyce, C. and Wood, A. 2011. Personality and the marginal utility of income: Personality interacts with increases in household income to determine life satisfaction. *Journal of Economic Behavior & Organization*. **78**, pp.183–191.

Breusch, T. S., & Pagan, A. R. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, **47**(5), 1287.

Brons, M. and Rietveld, P. 2009. Improving the quality of the door-to-door rail journey: A customer-oriented approach. *Built Environment*. **35**(1), pp.122–135.

Budzynski, T.H., Stoyva, J.M., Adler, C.S. and Mullaney, D.J. 1973. EMG biofeedback and tension headache: a controlled outcome study. *Psychosomatic medicine*. **35**(6), pp.484–496.

Calastri, C., Pawlak, J. and Batley, R. 2022. Participation in online activities while travelling: an application of the MDCEV model in the context of rail travel. *Transportation*. **49**(1), pp.61–87.

Campaign for Better Transport 2015. *Passenger's Guide to Franchising*.

Cao, X.J. and Ettema, D.F. 2014. Satisfaction with travel and residential self-selection: How do preferences moderate the impact of the hiawatha light rail. *Journal of Transport and Land Use*. **7**(3), pp.93–108.

Capuano, A. W., Dawson, J. D., & Gray, G. C. 2007. Maximizing power in seroepidemiological studies through the use of the proportional odds model. *Influenza and Other Respiratory Viruses*, **1**(3), 87–93.

Carrel, A., Mishalani, R.G., Sengupta, R. and Walker, J.L. 2016. In pursuit of the happy transit rider: Dissecting satisfaction using daily surveys and tracking data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. **20**(4), pp.345–362.

Cats, O., Abenoza, R.F., Liu, C. and Susilo, Y.O. 2015. Evolution of satisfaction with public transport and its determinants in Sweden identifying priority areas. *Transportation Research Record*. **2538**, pp.86–96.

České dráhy 2016. Odškodnění cestujících. [Accessed 10 February 2022]. Available from: https://www.cd.cz/typy-jizdenek/odskodneni-reklamace-a-vymeny/-26400/.

Ceyisakar, I. E., van Leeuwen, N., Dippel, D. W. J., Steyerberg, E. W., & Lingsma, H. F. 2021. Ordinal outcome analysis improves the detection of between-hospital differences in outcome. *BMC Medical Research Methodology*. **21**(1).

Chiou, Y.-C. and Chen, Y.-H. 2010. Factors influencing the intentions of passengers regarding full service and low cost carriers: A note. *Journal of Air Transport Management*. **16**, pp.226–228.

Choi, B. C., & Pak, A. W. 2005. A Catalog of Biases in Questionnaires. *Prev Chronic Dis.* ***2***(1).

Cimpean, A. and David, D. 2019. The mechanisms of pain tolerance and pain-related anxiety in acute pain. *Health Psychology Open*. **6**(2), p.2055102919865161.

Cipolletta, S., Andreghetti, G.R. and Mioni, G. 2022. Risk perception towards COVID-19: A systematic review and qualitative synthesis. *International Journal of Environmental Research and Public Health*. **19**(8).

Coker, J. and Izaret, J.M. 2021. Progressive pricing: The ethical case for price personalization. *Journal of Business Ethics*. **173**(2), pp.387–398.

Cooper, C.H.V. 2020. Quantitative models of well-being to inform policy: Problems and opportunities. *Sustainability*. **12**(8), pp.1–13.

Coppola, P. and De Fabiis, F. 2021. Impacts of interpersonal distancing on-board trains during the COVID-19 emergency. *European Transport Research Review*. **13**(1), p.13.

Cornet, Y., Lugano, G., Georgouli, C. and Milakis, D. 2022. Worthwhile travel time: a conceptual framework of the perceived value of enjoyment, productivity and fitness while travelling. *Transport Reviews*. **42**(5), pp.580–603.

DeCoster, J., Iselin, A. M. R., & Gallucci, M. 2009. A Conceptual and Empirical Examination of Justifications for Dichotomization. *Psychological Methods*, ***14***(4), 349–366. https://doi.org/10.1037/A0016956

Crombez, G., Van Damme, S. and Eccleston, C. 2005. Hypervigilance to pain: an experimental and clinical analysis. *Pain*. **116**(1–2), pp.4–7.

Daly, A., Tsang, F. and Rohr, C. 2014. The value of small time savings for non-business travel. *Journal of Transport Economics and Policy*. **48**(2), pp.205–218.

Dargay, J.M. 2002. Determinants of car ownership in rural and urban areas: A pseudo-panel analysis. *Transportation Research Part E: Logistics and Transportation Review*. **38**(5), pp.351–366.

Deole, S.S., Deter, M. and Huang, Y. 2023. Home sweet home: Working from home and employee performance during the COVID-19 pandemic in the UK. *Labour economics*. **80**, p.102295.

Department for Transport 2016. *Improving access to passenger compensation for delays and cancellations*.

Department for Transport 2020. *Rail Delays and Compensation 2020. Moving Britain Ahead.*

Department for Transport. 2021. *Great British Railways. The Williams-Shapps Plan for Rail.*

DeSerpa, A.C. 1971. A theory of the economics of time. *The Economic Journal*. **81**(324), p.828.

Dickerson, A., Hole, A.R. and Munford, L.A. 2014. The relationship between well-being and commuting revisited: Does the choice of methodology matter? *Regional Science and Urban Economics*. **49**, pp.321–329.

Dobruszkes, F. 2006. An analysis of European low-cost airlines and their networks. *Journal of Transport Geography*. **14**(4), pp.249–264.

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, **6**(3), pp.241–252.

Dziekan, K. and Kottenhoff, K. 2007. Dynamic at-stop real-time information displays for public transport: effects on customers. *Transportation Research Part A: Policy and Practice*. **41**(6), pp.489–501.

Eboli, L. and Mazzulla, G. 2021. Customer satisfaction as a measure of service quality in public transport planning. *International Encyclopedia of Transportation*. **6**, pp.220–224.

Efthymiou, M., Njoya, E.T., Lo, P.L., Papatheodorou, A. and Randall, D. 2019. The impact of delays on customers' satisfaction: An empirical analysis of the british airways on-time performance at heathrow airport. *Journal of Aerospace Technology and Management*. **11**, pp.1–13.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D. and Kruger, J. 2008. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*. **105**(1), pp.98–121.

ERA 2019. *An ERA study into regulation EU261: passenger compensation for delayed or* *cancelled flights*.

Ettema, D.F., Abenoza, R.F. and Susilo, Y.O. 2016. Satisfaction with intermodal trips in Stockholm: How do service attributes influence satisfaction with the main mode and with the journey as a whole? *Transportation Research Board 95th Annual Meeting. No. 16-2247*. **46**, pp.1–16.

Europe Economics 2019. *Delay Repay Claims Companies Market Review*. London.

European Commission 2007. *Regulation (EC) No 1371/2007 of the European Parliament and of the Council of 23 October 2007 on rail passengers' rights and obligations*.

European Commission 2004. *Regulation (EC) No 261/2004 of the European Parliament and of the Council of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, and repealin*. European Commission.

Evans, D.J. 2005. The elasticity of marginal utility of consumption: Estimates for 20 OECD countries. *Fiscal Studies*. **26**(2), pp.197–224.

Farrington, D. P., & Loeber, R. 2000. Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health*. **10**(2), 100–122.

Farzin, Y. 2009. The effect of non-pecuniary motivations on labor supply. *The Quarterly Review of Economics and Finance*. **49**, pp.1236–1259.

Ferrer-i-Carbonell, A. and Frijters, P. 2004. How important is methodology for the estimates of the determinants of happiness? *Economic Journal*. **114**(497), pp.641–659.

Ferrer-i-Carbonell, A. and van Praag, B.M.S. 2002. The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal. *Health Economics*. **11**(8), pp.709–722.

Fitzsimons, G. J. 2008. Death to Dichotomizing. *Journal of Consumer Research*, **35**(1), pp. 5–8.

Fleurbaey, M. 2009. Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*. **47**(4), pp.1029–1075.

Fosgerau, M. and Karlström, A. 2010. The value of reliability *In*: *Transportation Research Part B: Methodological*..pp.38–49.

Frank, L.D., Sallis, J.F., Saelens, B.E., Leary, L., Cain, L., Conway, T.L. and Hess, P.M.

2010. The development of a walkability index: application to the Neighborhood Quality of Life Study. *British Journal of Sports Medicine*. **44**(13), pp.924–933.

Frey, B.S. and Stutzer, A. 2009. *The life satisfaction approach to environmental valuation*.

Freyd, M. 1923. The graphic rating scale. *Journal of Educational Psychology*. **14**(2), pp.83–102.

Friman, M. and Fellesson, M. 2009. Service supply and customer satisfaction in public transportation: The quality paradox. *Journal of Public Transportation*. **12**(4), pp.57–69.

Gao, Y., Rasouli, S., Timmermans, H. and Wang, Y. 2018. Trip stage satisfaction of public transport users: A reference-based model incorporating trip attributes, perceived service quality, psychological disposition and difference tolerance. *Transportation Research Part A: Policy and Practice*. **118**, pp.759–775.

Gaver, D.P. 1968. Headstart strategies for combating congestion. *Transportation Science*. **2**(2), pp.172–181.

Giustinelli, P., Manski, C.F. and Molinari, F. 2019. Precise or imprecise probabilities? Evidence from survey response on late-onset dementia. *National Bureau of Economic Research Working Paper Series*. **No. 26125**.

GLA Economics 2005. *Time is money The economic effects of transport delays in Central London*.

Glazener, A., Sanchez, K., Ramani, T., Zietsman, J., Nieuwenhuijsen, M.J., Mindell, J.S., Fox, M. and Khreis, H. 2021. Fourteen pathways between urban transportation and health: A conceptual model and literature review. *Journal of Transport and Health*. **21**, p.101070.

Global Railway Review. 2022. *Chiltern Railways launch new Delay Repay compensation scheme*.

González-Roldán, A.M., Terrasa, J.L., Sitges, C., van der Meulen, M., Anton, F. and Montoya, P. 2020. Age-related changes in pain perception are associated with altered functional connectivity during resting state. *Frontiers in Aging Neuroscience*. **12**, p.116.

Gov.uk 2020. Rail passenger compensation paid by train operating companies. [Accessed 10 March 2022]. Available from: https://www.gov.uk/government/publications/train-operating-companies-passengers-charter-compensation/train-operating-companies-passengers-charter-compensation.

Great Western Railway 2019. *Annual Stakeholder Report 2018-19*.

Greene, W. H. 2019. *Econometric Analysis* (8th ed.). Pearson.

Greene, D.L., Kontou, E., Borlaug, B., Brooker, A. and Muratori, M. 2020. Public charging infrastructure for plug-in electric vehicles: What is it worth? *Transportation Research Part D: Transport and Environment*. **78**, p.102182.

Haefeli, M. and Elfering, A. 2006. Pain assessment. *European Spine Journal*. **15**, pp.S17-24.

Hardcastle, R. 2012. *How can we incentivise pension saving? A behavioural perspective.* Department for Work and Pension. Working paper 109.

Harris, J.K. 2021. Primer on binary logistic regression. *Family Medicine and Community Health*. **9**(Suppl 1), p.1290.

Haylen, A. 2019. *Rail passenger rights, compensation & complaints*. House of Commons Library..

Hess, S., Adler, T. and Polak, J.W. 2007. Modelling airport and airline choice behaviour with the use of stated preference survey data. *Transportation Research Part E: Logistics and Transportation Review*. **43**(3), pp.221–233.

Hess, S., & Palma, D. 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*. **32.** 100170.

Higgins, C.D., Sweet, M.N. and Kanaroglou, P.S. 2018. All minutes are not equal: travel time and the effects of congestion on commute satisfaction in Canadian cities. *Transportation*. **45**(5), pp.1249–1268.

Holmgren, J. 2013. An analysis of the determinants of local public transport demand focusing the effects of income changes. *European Transport Research Review*. **5**(2), pp.101–107.

Hörcher, D., Singh, R. and Graham, D.J. 2022. Social distancing in public transport: mobilising new technologies for demand management under the Covid-19 crisis. *Transportation*. **49**(2), pp.735–764.

Horowitz, J.L. and Savin, N.E. 2001. Binary response models: Logits, probits and semiparametrics. *Journal of Economic Perspectives*. **15**(4), pp.43–56.

Ilkley Gazette 2021. Trains cancelled between Ilkley and Leeds as emergency incident blocks line. [Accessed 10 November 2022]. Available from: https://www.ilkleygazette.co.uk/news/19576302.trains-cancelled-ilkley-leeds-emergency-incident-blocks-line/.

Institute for Transport Studies, Leigh Fisher, Rand Europe, & Systra. 2016. *Rail Demand Forecasting Estimation*.

Jensen, G.F., Erickson, M.L. and Gibbs, J.P. 1978. Perceived Risk of Punishment and Self-Reported Delinquency. *Social Forces*. **57**(1), p.57.

Joly, I. 2004. Travel Time Budget – Decomposition of the Worldwide Mean. *International Association of Time-Use ResearchAnnual Conference 200427-29 October, Rome Italy*.

Kahneman, D. and Krueger, A.B. 2006. Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*. **20**(1), pp.3–24.

Klemperer, P. 1995. Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies*. **62**(4), pp.515–539.

Knight, T.E. 1974. An approach to the evaluation of changes in travel unreliability: A ‘Safety margin’ hypothesis. *Transportation*. **3**(4), pp.393–408.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*. **47**(260), pp.583–621.

Lättman, K., Friman, M. and Olsson, L.E. 2016. Perceived accessibility of public transport as a potential indicator of social inclusion. *Social Inclusion*. **4**(3), pp.36–45.

Lau, J.T.F., Yang, X., Tsui, H. and Kim, J.H. 2003. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. *Journal of epidemiology and community health*. **57**(11), pp.864–870.

Layard, R. 2006. Happiness and public policy: A challenge to the profession. *Economic Journal*. **116**(510), pp.C24–C33.

Layard, R., Mayraz, G. and Nickell, S. 2008. The marginal utility of income. *Journal of Public Economics*. **92**(8), pp.1846–1857.

Leicester, A., Levell, P., & Rasul, I. 2012. *Tax and benefit policy: insights from* *behavioural economics*. Institute for Fiscal Studies (C125).

Lewis, A. and Duch, R. 2021. Gender differences in perceived risk of COVID-19. *Social Science Quarterly*. **102**(5), pp.2124–2133.

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**, pp.151–163.

Lochner, L. 2007. Individual perceptions of the criminal justice system. *American Economic Review*. **97**(1), pp.444–460.

Lora, E. 2016. The distance between perception and reality in the social domains of life. *Handbook of Happiness Research in Latin America*., pp.531–555.

Lunke, E.B. 2020. Commuters' satisfaction with public transport. *Journal of Transport and Health*. **16**, p.100842.

Lupyan, G. 2017. How reliable is perception? *Philosophical Topics*. **45**(1), pp.81–106.

Lyons, G., Jain, J. and Holley, D. 2007. The use of travel time by rail passengers in Great Britain. *Transportation Research Part A: Policy and Practice*. **41**(1), pp.107–120.

Lyons, G., Jain, J. and Weir, I. 2016. Changing times – A decade of empirical insight into the experience of rail passengers in Great Britain. *Journal of Transport Geography*. **57**, pp.94–104.

Lyons, G. and Urry, J. 2005. Travel time use in the information age. *Transportation Research Part A: Policy and Practice*. **39**(2), pp.257–276.

Machado-León, J.L., de Oña, R., Baouni, T. and de Oña, J. 2017. Railway transit services in Algiers: priority improvement actions based on users perceptions. *Transport Policy*. **53**, pp.175–185.

Mackerron, G. 2012. Happiness economics from 35000 feet. *Journal of Economic Surveys*. **26**(4), pp.705–735.

Mackie, P.J., Jara-Díaz, S. and Fowkes, A.S. 2001. The value of travel time savings in evaluation. *Transportation Research Part E: Logistics and Transportation Review*. **37**(2), pp.91–106.

Mackie, P.J., Wardman, M., Fowkes, A.S., Whelan, G., Nellthorp, J. and Bates, J. 2003. Values of Travel Time Savings UK. *Institute for Transport Studies*., p.123.

Mahmassani, H.S. and Chang, G.L. 1986. Experiments with departure time choice dynamics of urban commuters. *Transportation Research Part B: Methodological*. **20**(4), pp.297–320.

Mancini, F., Longo, M.R., Kammers, M.P.M. and Haggard, P. 2011. Visual distortion of body size modulates pain perception. *Psychological science*. **22**(3), pp.325–330.

Manning, M., Fleming, C.M., Pham, H.T. and Wong, G.T.W. 2022. What matters more, perceived or real crime? *Social Indicators Research*. **163**(3), pp.1221–1248.

Manor, O., Matthews, S., & Power, C. 2000. Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology*, **29**(1), 149–157.

Manski, C.F. 2004. Measuring expectations. *Econometrica*. **72**(5), pp.1329–1376.

Manski, C.F. and Molinari, F. 2010. Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics*. **28**(2), pp.219–231.

Marchetti, C. 1994. Anthropological invariants in travel behavior. *Technological Forecasting and Social Change*. **47**(1), pp.75–88.

Margolis, R.B., Tait, R.C. and Krause, S.J. 1986. A rating system for use with patient pain drawings. *Pain*. **24**(1), pp.57–65.

Marquardt, D. W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*. **12**(3), pp.591–612.

Masuda, Y.J., Williams, J.R. and Tallis, H. 2021. Does life satisfaction vary with time and income? Investigating the relationship among free time, income, and life satisfaction. *Journal of Happiness Studies*. **22**(5), pp.2051–2073.

McCullagh, P. 1980. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **42**(2), pp.109–127.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers of econometrics*. Academic Press.

McIntyre, M.H., Team, 23andMe Research, Kless, A., Hein, P., Field, M. and Tung, J.Y. 2020. Validity of the cold pressor test and pain sensitivity questionnaire via online self-administration. *PLOS ONE*. **15**(4), p.e0231697.

Mokhtarian, P.L. and Salomon, I. 2001. How derived is the demand for travel? Some conceptual and measurement considerations. *Transportation Research Part A: Policy and Practice*. **35**(8), pp.695–719.

Monsuur, F., Enoch, M., Quddus, M. and Meek, S. 2017. Impact of train and station types on perceived quality of rail service. *Transportation Research Record*. **2648**(1), pp.51–59.

Monsuur, F., Enoch, M., Quddus, M. and Meek, S. 2021. Modelling the impact of rail delays on passenger satisfaction. *Transportation Research Part A: Policy and Practice*. **152**, pp.19–35.

Mouwen, A. 2015. Drivers of customer satisfaction with public transport services. *Transportation Research Part A: Policy and Practice*. **78**, pp.1–20.

Nagy, E. and Csiszár, C. 2015. Analysis of delay causes in Railway passenger transportation. *Periodica Polytechnica Transportation Engineering*. **43**(2), pp.73–80.

Nash, C., Nilsson, J.E. and Link, H. 2013. Comparing three models for introduction of competition into railways. *Journal of Transport Economics and Policy*. **47**(PART2), pp.191–206.

Nathanail, E. 2008. Measuring the quality of service for passengers on the hellenic railways. *Transportation Research Part A: Policy and Practice*. **42**(1), pp.48–66.

Network Rail n.d. Railway performance. [Accessed 12 January 2021a]. Available from: https://www.networkrail.co.uk/who-we-are/how-we-work/performance/railway-performance/.

Network Rail 2012. *Schedule 8 compensation payment rates in CP5*.

Network Rail n.d. Technical overview: Payments relating to disruption. [Accessed 10 January 2020b]. Available from: https://cdn.networkrail.co.uk/wp-content/uploads/2019/03/Technicaloverview-%0Apayments-relating-to-disruption.pdf.

von Neumann, J. and Morgenstern, O. 2007. *Theory of games and economic behavior* 2nd ed. Princeton: Princeton University Press.

NEXTOR 2010. Total delay impact study : a comprehensive assessment of the costs and impacts of flight delay in the United States.

Nielsen, O.A. 2000. A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological*. **34**(5), pp.377–402.

Nielsen, O.A., Landex, A. and Frederiksen, R.D. 2009. Passenger delay models for rail networks. *Operations Research/ Computer Science Interfaces Series*. **46**(1), pp.27–49.

Noland, R.B. and Small, K.A. 1995. Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation Research Record*. **1493**(1493), pp.150–158.

O’Connell, J.F. and Williams, G. 2005. Passengers’ perceptions of low cost airlines and full service carriers: A case study involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines. *Journal of Air Transport Management*. **11**(4), pp.259–272.

Obsie, A., Woldeamanuel, M. and Woldetensae, B. 2020. Service quality of Addis Ababa light rail transit: Passengers’ views and perspectives. *Urban Rail Transit*. **6**(4), pp.231–243.

Ojeda-Cabral, M., Shires, J., Wardman, M., Teklu, F. and Harris, N. 2021. The use of recovery time in timetables: rail passengers’ preferences and valuation relative to travel time and delays. *Transportation*. **48**(1), pp.337–368.

Oliveira, L., Bruen, C., Birrell, S. and Cain, R. 2019. What passengers really want: Assessing the value of rail innovation to improve experiences. *Transportation Research Interdisciplinary Perspectives*. **1**, p.100014.

De Oña, J. and De Oña, R. 2015. Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. *Transportation Science*. **49**(3), pp.605–622.

De Oña, J., De Oña, R., Eboli, L., Forciniti, C. and Mazzulla, G. 2016. Transit passengers’ behavioural intentions: the influence of service quality and customer satisfaction. *Transportmetrica A Transport Science*. **12**(5), pp.385–412.

ORR 2016. *Office of Rail and Road super-complaint response report: Which? Super-complaint - compensation arrangements in the market for passenger rail services.*

ORR 2020. ORR Data Portal. [Accessed 10 March 2022]. Available from: https://dataportal.orr.gov.uk/.

ORR. 2021. *PR23 – Review of the Schedule 8 train performance regime. Technical Consultation.*

ORR 2014. *Passenger compensation and refund rights for delays and cancellations*.

Papamitsiou, Z. and Economides, A.A. 2014. *Students' perception of performance vs. actual performance during computer-based testing: a temporal approach*.

Papke, L. E., & Wooldridge, J. M. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*. **145**(1–2), pp. 121–133.

Parbo, J., Nielsen, O.A. and Prato, C.G. 2016. Passenger Perspectives in Railway Timetabling: A Literature Review. *Transport Reviews*. **36**(4), pp.500–526.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J. and White, P. 2006. The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*. **13**, pp.295–306.

Peer, S., Koopmans, C.C. and Verhoef, E.T. 2012. Prediction of travel time variability for cost-benefit analysis. *Transportation Research Part A: Policy and Practice*. **46**(1), pp.79–90.

Phelps, C.E. 2019. A new method to determine the optimal willingness to pay in cost-effectiveness analysis. *Value in Health*. **22**(7), pp.785–791.

Pollitt, M.G. and Smith, A.S.J. 2002. The restructuring and privatisation of British Rail: was it really that bad? *Fiscal Studies*. **23**(4), pp.463–502.

Pot, F., Wee, B. and Tillema, T. 2021. Perceived accessibility: What it is and why it differs from calculated accessibility measures based on spatial data. *Journal of Transport Geography*. **94**, p.103090.

Pratt, S. and Schuckert, M. 2018. Economic impact of low-cost carrier in a saturated transport market: Net benefits or zero-sum game? *Tourism Economics*. **25**(2), pp.149–170.

Preston, J. 2008. A review of passenger rail franchising in Britain: 1996/1997-2006/2007. *Research in Transportation Economics*. **22**(1), pp.71–77.

Preston, J. 2009. Trends in European railways over the last two decades. *Built Environment*. **35**(1), pp.11–23.

Preston, J., Wall, G., Batley, R., Ibáñez, J.N. and Shires, J. 2009. Impact of delays on passenger train services: Evidence from Great Britain. *Transportation Research Record*. **2117**, pp.14–23.

Rätzel, S. 2012. Labour supply, life satisfaction, and the (dis)utility of work. *Scandinavian*

Journal of Economics. **114**(4), pp.1160–1181.

RegioJet n.d. On Time Arrival Guarantee. [Accessed 10 February 2022]. Available from: https://regiojet.com/our-tickets/guarantee.

Renfe 2019. *Informe de responsabilidad social y gobierno corporativo*.

Rezapour, M. and Ferraro, F.R. 2021. Rail transport delay and its effects on the perceived importance of a real-time information. *Frontiers in Psychology*. **12**, p.619308.

Rich, J., Myhrmann, M.S. and Mabit, S.E. 2023. Our children cycle less - A Danish pseudo-panel analysis. *Journal of Transport Geography*. **106**.

Rietveld, P. 2002. Rounding of Arrival and Departure Times in Travel Surveys: An Interpretation in Terms of Scheduled Activities. *Tinbergen Institute. Working paper*. **01-110/3**, pp.71–82.

Rietveld, P., Bruinsma, F.R. and 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**(6), pp.539–559.

Rong, R., Liu, L., Jia, N. and Ma, S. 2022. Impact analysis of actual traveling performance on bus passenger's perception and satisfaction. *Transportation Research Part A: Policy and Practice*. **160**, pp.80–100.

Saelens, B.E., Sallis, J.F., Black, J.B. and Chen, D. 2003. Neighborhood-based differences in physical activity: an environment scale evaluation. *American journal of public health*. **93**(9), pp.1552–1558.

Sankey, S. S., Weissfeld, L. A., Sankey, S. S., & Weissfeld, L. A. 1998. A study of the effect of dichotomizing ordinal data upon modeling. *Communications in Statistics - Simulation and Computation*. **27**(4), pp.871–887.

Sanko, N. and Iriguchi, N. 2022. Are self-reported times rounded? Insights from times reported by an objective third party. *Transportation Research Interdisciplinary Perspectives*. **16**, p.100698.

Sato, Y. and Maruyama, T. 2020. Modeling the rounding of departure times in travel surveys: comparing the effect of trip purposes and travel modes. *Transportation Research Record*. **2674**(10), pp.628–637.

Schmid, B., Molloy, J., Peer, S., Jokubauskaite, S., Aschauer, F., Hössinger, R., Gerike, R., Jara-Diaz, S.R. and Axhausen, K.W. 2021. The value of travel time savings and the value of leisure in Zurich: Estimation, decomposition and policy implications. *Transportation Research Part A: Policy and Practice*. **150**, pp.186–215.

Schneider, C.R., Dryhurst, S., Kerr, J., Freeman, A.L.J., Recchia, G., Spiegelhalter, D. and van der Linden, S. 2021. COVID-19 risk perception: a longitudinal analysis of its predictors and associations with health protective behaviours in the United Kingdom. *Journal of Risk Research*. **24**(3–4), pp.294–313.

Shapiro, S. S., & Wilk, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, **52**(3–4), pp. 591–611.

Shelat, S., van de Wiel, T., Molin, E., van Lint, J.W.C. and Cats, O. 2022. Analysing the impact of COVID-19 risk perceptions on route choice behaviour in train networks. *PLOS ONE*. **17**(3), p.e0264805.

Shepperd, J.A., Klein, W.M.P., Waters, E.A. and Weinstein, N.D. 2013. Taking stock of unrealistic optimism. *Perspectives on Psychological Science*. **8**(4), pp.395–411.

Small, K.A. 1982. The scheduling of consumer activities: work trips. *American Economic Review*. **72**(3), pp.467–479.

Smith, A.S.J. and Ojeda Cabral, M. 2022. Is higher quality always costly? Marginal costs of quality: Theory and application to railway punctuality. *Transportation Research Part A: Policy and Practice*. **157**, pp.258–273.

Soza-Parra, J., Raveau, S., Muñoz, J.C. and Cats, O. 2019. The underlying effect of public transport reliability on users' satisfaction. *Transportation Research Part A: Policy and Practice*. **126**, pp.83–93.

Spearman, C. 1904. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*. **15**(1), pp.72.

Spicer, V., Song, J. and Brantingham, P. 2014. Bridging the perceptual gap: variations in crime perception of businesses at the neighborhood level. *Security Informatics*. **3**(1), p.14.

St-Louis, E., Manaugh, K., van Lierop, D. and El-Geneidy, A. 2014. The happy commuter: A comparison of commuter satisfaction across modes. *Transportation Research Part F: Traffic Psychology and Behaviour*. **26**, pp.160–170.

StataCorp 2021. Stata Statistical Software: Release 17.

Stead, A.D., Wheat, P., Smith, A.S.J.J. and Ojeda-Cabral, M. 2019. Competition for and in the passenger rail market: Comparing open access versus franchised train operators' costs and reliability in Britain. *Journal of Rail Transport Planning & Management*. **12**, p.100142.

Steer 2018. *Schedule 8 - National Recalibration Methodology (Control Period 6)*.

Stevens, J.P. 2001. *Applied Multivariate Statistics for the Social Sciences*. Routledge.

Stutzer, A. and Frey, B.S. 2008. Stress that doesn't pay: The commuting paradox. *Scandinavian Journal of Economics*. **110**(2), pp.339–366.

Susilo, Y.O. and Cats, O. 2014. Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. *Transportation Research Part A: Policy and Practice*. **67**, pp.366–380.

The Independent 2018. Ryanair court case shows why flight delay compensation is fundamentally flawed. [Accessed 15 November 2019]. Available from: https://www.independent.co.uk/travel/news-and-advice/ryanair-flight-delay-compensation-overbooking-claims-flawed-court-case-ruling-a8268886.html.

The Local 2019. Rail passengers in germany paid €53.6 million compensation over late trains. [Accessed 10 February 2022]. Available from: https://www.thelocal.de/20190219/train-passengers-in-germany-given/.

The Social Market Foundation 2015. *Should switch, don't switch. Overcoming consumer inertia*.

Tiznado-Aitken, I., Lucas, K., Muñoz, J.C. and Hurtubia, R. 2020. Understanding accessibility through public transport users' experiences: A mixed methods approach. *Journal of Transport Geography*. **88**, p.102857.

Transport Focus 2019. *National Rail Passenger Survey. Main Report: Autumn 2019*.

Transport Focus. 2020. *National Rail Passenger Survey. Technical Report. Spring 2020 (Wave 42)*.

Transport Focus 2020a. *National Rail Passenger Survey. Main Report: Spring 2020*.

Transport Focus 2020b. *NRPS Spring 2020. Quality Assurance Statement.*

Transport Focus 2015. *Train punctuality: The passenger perspective*.

Tsoleridis, P., Choudhury, C.F. and Hess, S. 2022. Deriving transport appraisal values from emerging revealed preference data. *Transportation Research Part A: Policy and Practice*. **165**, pp.225–245.

Týfa, L., Jacura, M., Svetlík, M. and Vachtl, M. 2010. Acceptance of train delays by passengers. *Transactions on Transport Sciences*. **3**(2), pp.83–92.

Vallejo Velazquez, M., Kounadi, O. and Podor, A. 2020. Analysis and mapping of crime perception: A quantitative approach of sketch maps. *AGILE: GIScience Series*. **1**, pp.1–18.

Vansteenwegen, P. and Van Oudheusden, D. 2007. Decreasing the passenger waiting time for an intercity rail network. *Transportation Research Part B: Methodological*. **41**(4), pp.478–492.

Verma, A., Hanna, M., Lewin, V., Robinson, B. and Robinson, R. 2015. PTU-019 Pain, perception and reality – comparison of comfort scores of south asian and white british patients undergoing bowel cancer screening colonoscopy. *Gut*. **64**(Suppl 1), pp.A66–A66.

Vickerman, R. 2021. Will Covid-19 put the public back in public transport? A UK perspective. *Transport policy*. **103**, pp.95–102.

Vickrey, W.S. 1969. Congestion theory and transport investment. *The American Economic Review*. **59**(2), pp.251–260.

De Vos, J., Lättman, K., van der Vlugt, A.-L., Welsch, J. and Otsuka, N. 2023. Determinants and effects of perceived walkability: a literature review, conceptual model and research agenda. *Transport Reviews*. **43**(2), pp.303–324.

De Vos, J., Mokhtarian, P.L., Schwanen, T., Van Acker, V. and Witlox, F. 2016. Travel mode choice and travel satisfaction: bridging the gap between decision utility and experienced utility. *Transportation*. **43**(5), pp.771–796.

De Vos, J., Schwanen, T., van Acker, V. and Witlox, F. 2013. Travel and subjective well-being: A focus on findings, methods and future research needs. *Transport Reviews*. **33**(4), pp.421–442.

Wan, D., Kamga, C., Hao, W., Sugiura, A. and Beaton, E.B. 2016. Customer satisfaction with bus rapid transit: a study of New York City select bus service applying structural equation modeling. *Public Transport*. **8**(3), pp.497–520.

Wardman, M. 1998. A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy*. **22**, pp.71–91.

Wardman, M. 2001. A review of British evidence on time and service quality valuations. *Transportation Research Part E: Logistics and Transportation Review*. **37**(2–3), pp.107–128.

Wardman, M. and Batley, R. 2022. The demand impacts of train punctuality in great britain: systematic review, meta-analysis and some new econometric insights. *Transportation*. **49**(2), pp.555–589.

Wardman, M. and 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*. **41**(5), pp.1041–1069.

Wardman, M. and Lyons, G. 2016. The digital revolution and worthwhile use of travel time: implications for appraisal and forecasting. *Transportation*. **43**(3), pp.507–530.

Wardman, M. and Toner, J. 2020. Is generalised cost justified in travel demand analysis? *Transportation*. **47**(1), pp.75–108.

Welch, M. and Williams, H. 1997. The sensitivity of transport investment benefits to the evaluation of small travel-time savings. *Journal of Transport Economics and Policy*. **31**(3), pp.231–254.

Wheat, P. and Wardman, M. 2017. Effects of timetable related service quality on rail demand. *Transportation Research Part A: Policy and Practice*. **95**, pp.96–108.

Wilks, S. S. 1938. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, **9**(1), pp.60–62.

Wilson, C.M. and Price, C.W. 2005. *Irrationality in Consumers ' Switching Decisions : When More Firms May Mean Less Benefit* [Online]. University Library of Munich, Germany. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.163&rep=rep1&type=pdf.

Wittmer, A. and Laesser, C. 2010. The perception of time in air transport – What a delay is accepted by air travellers? *Journal of Air Transport Studies*. **1**(1), pp.48–61.

Yang, M., Zhao, J., Wang, W., Liu, Z. and Li, Z. 2015. Metro commuters' satisfaction in multi-type access and egress transferring groups. *Transportation Research Part D: Transport and Environment*. **34**, pp.179–194.

Yap, M. and Cats, O. 2021. Predicting disruptions and their passenger delay impacts for public transport stops. *Transportation*. **48**(4), pp.1703–1731.

Ye, R., De Vos, J. and Ma, L. 2022. New insights in travel satisfaction research.

Transportation Research Part D: Transport and Environment. **102**.

Zahavi, Y. 1974. Traveltime Budgets and Mobility in Urban Areas. , p.96.

Zahavi, Y. and Talvitie, A. 1980. Regularities in Travel Time and Money Expenditures. *Transportation Research Record*. (750), pp.13–19.