The use of recovery time in timetables: rail passengers' preferences and valuation relative to travel time and delays

Manuel Ojeda-Cabral[1] · Jeremy Shires[1] · Mark Wardman[1] · Fitsum Teklu[2] · Nigel Harris[3]

Abstract

Recovery time in the rail industry is the additional time that is included in train timetables over and above the minimum journey time necessary often with the explicit aim of improving punctuality. Recovery time is widely used in railways in a number of countries but prior to this study there has been no investigation of the rail users' point of view. Perceived recovery time, such as being held outside stations and prolonged stops at stations, might have some premium valuation due to the frustration caused. If perceived recovery time in train timetables does carry a premium, then the benefits of improved punctuality achieved by it will be reduced. This paper is the first to investigate passengers' views and preferences on the use of recovery time. We summarise the findings of a large study and provide estimates of passengers' valuations of recovery time, both relative to in-vehicle time and late time, that can be used for economic appraisal purposes. Overall, we find most passengers support the use of recovery time but the context is important. Only 13% of users disapprove of its use as a tool to reduce lateness. The estimated premia vary by demand characteristics and are significant in some contexts, although on average are of a small magnitude. The applicability of the estimates is demonstrated through the appraisal of an actual scheme in the UK. We observe that the introduction of more recovery time along with the subsequent improvement in reliability can lead to significant reductions in generalised journey time, even when recovery time carries a valuation premium. We must however strike a word of caution since we note that there were higher than expected proportions of non-traders in the survey which may have affected the results; future studies into the topic should look to minimise the proportion of non-traders. This study provides valuable and necessary first steps in this challenging topic.

Keywords Recovery time · Buffer time · Reliability valuation · Late time multiplier · Delay time · Railway timetable · Rail demand

✉ Manuel Ojeda-Cabral
  M.A.OjedaCabral@leeds.ac.uk

1 Institute for Transport Studies, University of Leeds, 30-40 University Road, Leeds LS2 9JT, UK

2 SYSTRA, 3rd Floor 5 Old Bailey, London, England EC4M 7BA, UK

3 The Railway Consultancy Ltd, 1st Floor South Tower, Crystal Palace Station, London SE19 2AZ, UK

 Springer

Introduction

Context

What is termed recovery time in the rail industry is the extra time or 'buffer' time that is included in train timetables over and above the minimum journey time necessary often with the explicit aim of improving punctuality. In addition to improved punctuality, it can manifest in a number of ways in cases where there is not much delay requiring the train to 'call on' the recovery time:

- Longer than necessary stops at intermediate stations;
- Slower than line-speed running between stations, which might involve trains being held prior to arriving in a station;
- Arriving earlier than scheduled.

By way of example, recovery time in Great Britain can be significant on some routes, extending journey times for longer distance travellers by even around 10% as recovery time is incurred through the course of a journey, whilst the insertion of 5 to 10 min extra time on long distance trains prior to the final destination can lead to very large proportionate increases in travel time for those using the service for shorter journeys.[1] Nor is recovery time limited to longer distance operators; for example, in 2004 South West Trains, serving suburban as well as longer distance routes from London Waterloo, undertook a major re-cast of its timetables with a central feature being additional recovery time to achieve a more robust timetable delivery. It is important to note that recovery times are not immediately apparent to passengers who only see the public timetable that includes it, even when they are 'following' their train live on a travel app. Only the train planners have information on the amount of recovery time.

Recovery time is not unique to railways and is also present in bus and airline schedules whilst motorists, cyclists and those walking can allow extra time to ensure a punctual arrival. In Great Britain, it has been included in railway timetables to improve punctuality, particularly as in recent years train operators and the infrastructure manager are liable to pay passengers and other parties monies to compensate for lost revenue due to delays. Reliability is also the most important driver of passenger satisfaction with the railways (Passenger Focus 2012). On the other hand, there is a push for timetables to deliver quicker journey times, and inserting recovery time would mean that the timetables are not optimal when there are no delays. There is a need to understand what the balance should be between the efficiency a quicker timetable delivers and the reliability a slower timetable that includes recovery time delivers. This is the subject of this paper.

Aims and structure

The impetus to this research was that several train operators in Great Britain felt that they might have inserted too much recovery time into timetables and were concerned that perceived recovery time, such as being held outside stations and prolonged stops at stations, might have some premium valuation due to the frustration caused. After all, waiting time

---

[1] In our understanding, and depending upon the level of congestion and relative difficulties in pathing, the total amount of recovery time added varies by route, but ranges broadly from 5 to 10%.

might be spent in conditions that are the same as (or approximate to) in-vehicle time yet the convention in transport planning practice worldwide is to value wait time at twice in-vehicle time (OECD/ITF 2014). If perceived recovery time in train timetables does have a premium valuation then the benefits of improved reliability achieved by it will be reduced.

The objective of this research was to understand rail passengers' views on this matter and estimate their valuations of changes in recovery time relative to in-vehicle time and reliability in the form of late time. This paper summarises the findings of a large study and provides estimates of passengers' valuations of recovery time, both relative to in-vehicle time and late time, that can be used for economic appraisal purposes. The fresh empirical evidence reported here provides an important contribution given that, as far as we were aware, there had been no prior research into passengers' preferences on recovery time. Previous research concerning recovery time has largely focussed on the supply side (e.g. UIC 2000; Schittenhelm 2011).

Given the complex issues involved and little prior research to guide us, the study was conducted in three phases: (1) an initial online survey of 1006 respondents, (2) two focus groups aimed at providing in-depth insights and informing the main survey, and (3) an on-train survey, providing information on the perceptions of and attitudes towards recovery time alongside the SP exercise used to examine preferences around recovery time.

"Background" section provides additional background on recovery time and briefly reviews the available literature on it. "Initial insights on passengers' perspectives on recovery time from surveys" section presents the key initial insights into passengers' perspectives of recovery time from the surveys and focus groups conducted in the study. The main quantitative phase of the study was based upon a Stated Preference (SP) valuation exercise as part of the final survey: "The stated preference experiment" section sets out how this was designed and a description of the data. "Analysis of stated preference data" reports the findings of the analysis of the SP data followed by an illustration of the use of values in an appraisal in "Illustrative use of values in appraisal" section. Concluding remarks are contained in "Conclusions".

Background

Recovery time in railway timetables

The published train timetables, upon which travellers make their decisions and plan their journeys, have always contained recovery time additional to the minimum necessary journey time, and there is a strong recognition of the role of recovery time in regulating travel time reliability (Parbo et al. 2016). Recovery time has traditionally been inserted into timetables for a number of reasons:

- Variability in traction performance. From the earliest days, some locomotives (even within the same batch) were more powerful or performed better;
- Variability in driver and signaller performance;
- Avoidance of conflicts at key junctions and points of congestion on the rail network;
- Organising train pathing, by providing the right sequence of trains and even, on occasions, improving connections by holding back a stopping service;
- Supporting a regular interval or clock-faced timetable, with Switzerland being a prime example;

- Service regulation and regularisation, on the grounds it is better to manage delays and out-of-course running before reaching a pinch-point;
- Dealing with variability in the volume of rail passengers and station dwell time;
- Allowing resilience for engineering based speed restrictions.

The over-arching advice on how to calculate recovery time is published by the International Union of Railways (UIC 2000). However, it is ultimately down to individual railway administrations and operators to calculate guidance on their own recovery margins (Palmqvist et al. 2017) and hence these differ between and within countries, with different route densities, rolling stock and route lengths all exerting an influence. We might expect the significance of recovery time to vary across countries if only because what is deemed to be acceptable lateness, termed a delay threshold, itself varies across countries (Li et al. 2010). For example, delay thresholds are set at 3 min for long distance trains in the Netherlands, 5 min in Denmark and Switzerland, through to 10 min in the UK and 15 min in Italy (Schittenhelm 2011). Typically, recovery time margins expressed as a percentage of nominal running time will vary between 3 and 7% in Europe and 6 to 8% in North America (Pachl 2002), but are driven primarily by operational rather than commercial considerations. Recovery time for commercial reasons, which is the purpose of this paper, adds to these non-trivial amounts of recovery time.

Regulatory and commercial influences

In recent years, there have been other incentives to improve reliability and the use of recovery time. In part, there has been increased recognition of the commercial implications of train reliability. But another raft of incentives, particularly in Great Britain, stem from the regulatory framework.

The 'Citizen's Charter' was a British political initiative launched by the government in 1991 with the aim of improving public services in the UK and making them more responsive to users. As far as the railways were concerned, it introduced the concept of passengers being compensated for unreliable services, and in the first instance was restricted to season ticket holders. The post-1996 privatisation form of this is the Passenger's Charter, which is a part of each train operating companies' franchise commitment. This sets out the conditions under which train operators pay compensation to passengers in the event of late arrivals. For example, on long-distance routes, if there is a delay on one leg of a return journey, passengers can expect reimbursement of ¼ of the value of the return ticket if they are 30 min late, ½ the value if an hour late, or all of it if 2 h late. More recently, urban operators such as c2c have been offering those paying with Smart cards very minor refunds (automatically paid electronically) starting if trains are only 3 min late. Such compensation procedures inevitably incentivise recovery time.

Parallel to this, the privatised railway industry in Britain separated operations and infrastructure and hence a regulatory system was introduced, termed 'Schedule 8', to incentivise good performance. It specified compensation rates to train operators for service disruptions caused by the infrastructure provider or by other train operators and sets out rewards to the infrastructure provider for better than target performance. Network Rail (NR), as the infrastructure provider in Britain, has incurred Schedule 8 payments to operators of £138

million in 2012/2013, £194 million in 2013/2014, £109 million in 2014/2015 and £106 million in 2015/2016,[2] denoting this is not a trivial issue.

Schedule 8 regulations do impact on the construction of timetables, even if in a subliminal fashion. Train operators may consider reducing recovery time, in order to trigger a larger number and greater amount of Schedule 8 payments, especially if they feel that passengers would not stop using the railway in large enough numbers in response to reduced reliability. In contrast, Network Rail might be incentivised to seek increased recovery time to reduce its exposure to compensation payments and indeed to receive bonus payments. However, because excessive recovery time in timetables also causes problems on a capacity-constrained railway (if trains arrive early, there may be nowhere to put them), a balance between these factors is found between NR and operators, even if only iteratively over a number of timetables. Nevertheless, variations in Schedule 8 payments have caused significant budgeting problems for train operators, for instance on the East Coast Main Line.

Previous research

As far as we are aware, there have been no previous studies of whether and to what extent rail travellers perceive and value the dwell time consequences of recovery time to be different to in-vehicle time. What we think is the closest proxy for this in the literature, because it also involves travelling at less than the 'normal' speed, is the congestion multiplier for car travel and the 'slowed-down' and 'dwell time' multiplier for bus travel.

There is now a wealth of evidence regarding the congested travel time multiplier for motorists. Wardman and Ibáñez (2012) provide an extensive summary of international evidence which suggests that a central value is 1.5. A more recent review (Wardman et al. 2016), covering a very large amount of European wide evidence, also returned a multiplier of around 1.5.

The only study of which we are aware that examined the valuation of slowed-down time and dwell time for bus is the third UK national value of time study (Batley et al. 2017). It reports a multiplier of around 1.4 for slowed down time for commuters and leisure travellers with leisure travellers having a multiplier of 1.6 for dwell time at the bus stop.

The purpose of recovery time is to improve travel time reliability, and valuations of the latter are a feature of the new evidence we here report. In contrast to values of rail recovery time, there is now an extensive literature covering valuations of travel time variability across all modes. The rail industry in Great Britain uniquely has its Passenger Demand Forecasting Handbook (PDFH) and this has recommended reliability values since its inception in 1986. The measure used is mean late time and currently recommended multipliers are in the range 2.3 to 3.9 depending upon flow type. Other significant reviews of rail reliability values are provided by Wardman and Batley (2014) in the UK context, Wardman et al. (2016) in a broader European context and OECD/ITF (2014) in an international context.

Initial insights on passengers' perspectives on recovery time from surveys

This section discusses the findings from a set of questions regarding passengers' awareness of and attitudes towards recovery time from the online and on-train surveys, as well as the focus groups where passengers provided further insights, particularly into the format of the SP exercise.

The online and on-train surveys

The online survey was undertaken in February 2012 and was completed by 1006 rail users from a national panel. The on-train survey was conducted in June 2012 on a mix of services. The large sample of 1013 passengers obtained was reduced slightly to 972 after removing those who did not provide all relevant details.

Table 1 allows a comparison of the features of the achieved samples with the National Travel Survey (NTS) which provides a representative account of travel in Great Britain obtained from a random sample of households. We took the latter to cover the years 2010 to 2015 and those aged 16 and over making rail trips which yields a sample of 8254 individuals. We report the proportion in each category along with its standard error. The NTS figures are weighted by the number of trips made by each respondent.

The journey purpose splits for the online survey result from quotas specified for the recruitment whilst the on-train surveys were conducted on a mix of long distance inter-city (East Coast Main Line and Cross Country), regional (TransPennine) and commuter (First Capital Connect and South West Trains) services as a practical means of surveying a range of routes and traveller characteristics given the resources available. Whilst there are some inevitable discrepancies between our samples and the NTS in terms of journey purpose, this is not a particular cause for concern since the descriptive statistics covering recovery time and the modelling of the SP data both stratify by journey purpose.

Nonetheless, when we examine the gender, age, employment status and occupation characteristics of the on-train and online samples, we find them to be encouragingly similar to the distributions in the NTS; this is particularly the case for the On-Train survey. Comparing NTS and On-Train survey, the youngest group (16–25) is slightly underrepresented and those above 45 years-old are also slightly over-represented; however, the figures for other age groups, gender, employment status and occupation categories are not significantly different from the NTS corresponding values. Thus we conclude that the On-Train sample, which will be the one used to derive valuation estimates, can be deemed acceptable on representativeness.

Awareness of the existence of recovery time

The online survey asked respondents if they had noticed a range of journey time 'irregularities' during their reported most recent journey. 15% stated that they had noticed trains travelling slower than normal between stations, 11% felt that their trains stopped for longer than necessary at intermediate stations and 23% thought their trains stopped or slowed down unexpectedly between stations. Out of all travellers that had experienced any of these

Table 1 Comparison of on-train and online samples with NTS

	NTS	On-train	Online
Journey purpose			
Commuting	0.25 (0.0045)	0.31 (0.015)	0.20 (0.013)
Business urban	0.03 (0.0018)	0.06 (0.008)	0.07 (0.008)
Business inter-urban	0.08 (0.0028)	0.28 (0.014)	0.33 (0.015)
Leisure urban	0.31 (0.0048)	0.10 (0.010)	0.09 (0.009)
Leisure inter-urban	0.33 (0.0049)	0.25 (0.014)	0.31 (0.014)
Gender			
Male	0.55 (0.0091)	0.57 (0.016)	0.45 (0.018)
Age group			
16–25	0.20 (0.0060)	0.12 (0.010)	0.13 (0.011)
26–34	0.22 (0.0066)	0.20 (0.013)	0.18 (0.012)
35–44	0.21 (0.0065)	0.22 (0.014)	0.20 (0.013)
45–54	0.18 (0.0057)	0.23 (0.014)	0.22 (0.013)
55–64	0.12 (0.0043)	0.16 (0.011)	0.15 (0.011)
65+	0.07 (0.0028)	0.07 (0.008)	0.12 (0.010)
Employment status			
Employed full time	0.67 (0.0101)	0.71 (0.015)	n.a.
Employed part time	0.12 (0.0043)	0.12 (0.010)	n.a.
Student	0.07 (0.0033)	0.06 (0.008)	n.a.
Not working	0.05 (0.0025)	0.02 (0.004)	n.a.
Retired	0.09 (0.0028)	0.09 (0.009)	n.a.
Occupation type			
Professional/senior managerial	0.73 (0.0095)	0.50 (0.016)	0.41 (0.015)
Middle managerial		0.20 (0.013)	
Junior Managerial/clerical/supervisory	0.07 (0.0034)	0.10 (0.010)	0.32 (0.015)
Skilled manual	0.08 (0.0038)	0.07 (0.008)	0.10 (0.009)
Unskilled manual	0.01 (0.0012)	0.02 (0.004)	0.17 (0.012)
No employed occupation	0.11 (0.0042)	0.11 (0.010)	

The online sample did not ask about employment status. The occupation types are driven by the on-train survey categories and these do not correspond exactly with those in the NTS and the online survey

circumstances, 74% stated that they were not told about the reasons causing them, which could be because they were planned.

Most people are unlikely to be familiar with the term 'recovery time', although the concept is straightforward. For this reason, respondents were informed that "train operators sometimes include in their timetables additional time over the minimum required to get to the destination to allow for unforeseen delays" and were then asked whether they were aware of this. As is apparent in Table 2, only 25% were aware in the online survey, with minor variations among the different user groups.

The proportion of aware passengers was somewhat larger (43%) in the on-train survey, being higher for commuters and lower for leisure travellers, and presumably this is because the on-train sample would contain more frequent rail travellers, who may have noticed different journey times by direction of travel. Moreover, the on-train surveys had a greater focus on routes where recovery time had been introduced. We would expect these

Table 2 Awareness of recovery time by journey purpose and distance

Awareness of recovery time	All (%)	Purpose and distance					
		Commute		Business		Leisure	
		Under 20 miles (%)	Over 20 miles (%)	Under 50 miles (%)	Over 50 miles (%)	Under 50 miles (%)	Over 50 miles (%)
% Aware online	25	26	27	29	29	21	23
%Aware on-train	43	47	47	42	42	39	39

proportions to grow over time both as recovery time becomes a more common feature of timetables and travellers become more aware of the practice. Moreover, the consequences of recovery time are additional waiting time at or between stations and travellers will experience the annoyance of this even if they are not aware of the presence of recovery time in timetables.

Finally, while a lower share of awareness could be attributed to lower frequency, the answers to subsequent questions made us believe that the online responses were somewhat less reliable than those from the on-train survey. This is not surprising given well-known limitations of online panels (see e.g. Significance et al. 2012). Thus, in what follows we focus on the figures from the on-train survey, since the on-train survey is the one actually carrying the valuation experiment.[3]

Awareness of the amount of recovery time used by train operators

Those who were aware of the concept of recovery time in timetables were asked "how much additional time do you think train operators allow for trains similar to the one you are currently on". Table 3 summarises the responses by purpose and distance band for the on-train survey.

The category 'cannot say' in most cases contains the largest proportion of respondents, highlighting the difficulty in this task. Where respondents could provide an estimate, by far the largest proportions are for 2–5 min. The second most common category tends to be 6–10 min, especially for the longer journeys as might be expected. Very few thought that recovery time was less than 2 min and it is only amongst business and leisure travellers on longer distance journeys where there is a noticeable proportion stating the recovery time exceeds 10 min. Perceived recovery time unsurprisingly tends to be larger for longer distance journeys but with no strong variations by journey purpose.

A follow-up question revealed that, of those aware of recovery time, 20% felt that it improves punctuality a lot, a further 50% stated that it improved punctuality a little with 19% feeling it made no difference and the remainder unable to say.

---

[3] More details from the online survey can be found in Teklu et al. (2012); in general findings are not too dissimilar but the online survey revealed a few unreasonable values e.g. for expected and ideal recovery time.

Table 3 Passengers' estimates of typical recovery time (column percentages)

Recovery time typically included by operators	All (%)	Commute		Business		Leisure	
		Under 30 min (%)	Over 30 min (%)	Under 90 min (%)	Over 90 min (%)	Under 90 min (%)	Over 90 min (%)
Cannot say	39	48	37	32	38	40	46
<2 min	6	5	11	7	2	6	2
2–5 min	33	43	36	49	29	42	19
6–10 min	15	5	13	10	18	11	18
11–15 min	5	0	2	1	10	0	11
16–30 min	2	0	2	1	3	1	4
>30 min	0	0	0	0	0	0	0

Table 4 Ideal rail recovery time (column percentages)

Ideal recovery time	Scheduled train time				
	Under 30 min (%)	31–60 min (%)	Between 1 and 2 h (%)	More than 2 h (%)	All (%)
Cannot say	18	15	20	14	16
<2 min	34	22	17	6	17
2–5 min	39	53	45	35	45
6–10 min	9	9	14	26	16
11–15 min	0	1	4	14	5
16–30 min	0	0	1	4	1
>30 min	0	0	0	0	0

Ideal recovery time

Obvious questions to ask of our samples are whether and to what extent recovery time is wanted. The on-train survey asked about the approval of recovery time in train timetables and about an ideal rail recovery time. Given contingency had been defined to respondents,[4] the question took the form "How much contingency would you ideally like to have on trains similar to the one you are currently on?". Regarding the 'approval' question 56% approved, 12% disapproved, and the remainder had no preference. More importantly, respondents were asked to select the recovery time they would like the train operators to add. The responses are summarised in Table 4. The responses reveal a clear correlation between journey length and ideal recovery time, as might be expected. The largest category chose between 2 and 5 min with two-thirds between 2 and 15 min. A non-negligible 16%

---

[4] During the focus groups (prior to the on-train surveys), after introducing the concept of recovery time, we explored the best terminology to use in our questionnaire. Terms in addition to recovery time were contingency time, extra time, catch-up time, buffer time, reliability time, allowance time, leeway, just in case time and emergency time, along with some others. Although not everyone preferred contingency time, it was understood by all and hence adopted in the questionnaire and SP exercise.

could not say, perhaps suggesting that optimal levels are also dependent on other factors such as reliability levels.

In summary, the surveys reveal that travellers' awareness of and preferences towards recovery time are mixed but that despite the limited awareness, a majority approves of recovery time as a tool to improve reliability. There is broad support for modest amounts of recovery time, prior to testing how individuals would trade-off recovery time with in-vehicle time and delays.

The focus groups

Two focus groups were undertaken, in Basingstoke and Birmingham, with the particular aim of testing the viability of an SP exercise and determining its most appropriate format, and guiding the developments of the questionnaire.

Whilst some of the focus group recruits were aware of the concept of recovery time, none were aware of the term 'recovery time'. After being informed of the concept, most thought that recovery time would improve punctuality, and perceived it in a positive way, although a small minority considered that train operators should not need recovery time in order to be punctual and considered it to be a negative feature. All but one of the 20 participants build in their own 'recovery' time for car journeys, but only a minority do so for train journeys and then it tends to be less than for car journeys. The reasoning for building in more car recovery time was because it is their own responsibility to arrive at the destination on time whereas for train journeys it was more the responsibility of the train operator who could be blamed or provide compensation. Participants were more likely to build in significant recovery time into their rail journeys when travelling to airports or when on business trips.

The rationale behind the SP exercise is set out below, but it involved specifying journeys over 5 typical days and for each day conveying different actual journey times, levels of punctuality and actual recovery time for the two options offered. The focus groups presented SP 'mock-ups' and these established that respondents could relate to the presentation of different journeys across 5 days, which is critical given that reliability is an inherent feature of the exercise. This was re-assuring given this now seems to be the accepted means of presentation in reliability studies. Nonetheless, we here have the added dimension of recovery time.

In one SP version, we provided the timetabled departure and arrival time, but participants felt this was an unnecessary level of detail and indeed the key information can be conveyed without this. A simple version was to present the scheduled time and the planned recovery time, and for each day the actual recovery time and the arrival time punctuality. Participants though preferred a version which additionally included the actual journey time which then made clear the levels of punctuality and actual recovery time. This was preferred to an alternative which instead of providing the actual journey time specified the en-route delay time on each day. Participants generally felt that providing both the en-route delay time and the actual journey time alongside the punctuality and actual recovery time on each day provided too much information to assimilate.

We therefore opted, as is apparent in Fig. 1 below, to provide the actual journey time on each day and hence the implied actual recovery time and the punctuality given the scheduled journey time and specified contingency.

Scenario 1	Train A		□₁	Train B		□₂
Scheduled Journey Time	1 hr 10 mins (15 minutes contingency)			50 mins (no contingency)		
Days in a typical week	Actual Journey Time	Wait time between a pair of stations	Punctuality	Actual Journey Time	Wait time between a pair of stations	Punctuality
Day 1	55 mins	15 mins	ON TIME	50 mins	none	ON TIME
Day 2	55 mins	15 mins	ON TIME	1 hr	none	10 min LATE
Day 3	1 hr 5 mins	5 mins	ON TIME	1 hr	none	10 min LATE
Day 4	1 hr 15 mins	0 mins	5 mins LATE	1 hr 10 mins	none	20 min LATE
Day 5	1 hr 15 mins	0 mins	5 mins LATE	1 hr 20 mins	none	30 min LATE

Fig. 1 Example of SP choice scenario

The stated preference experiment

Design of the stated preference experiments

In order to infer valuations of recovery time, meaningful trade-offs had to be offered to travellers. We could not envisage a context where we could simply offer trade-offs between recovery time and journey time in a realistic manner. Since the purpose of recovery time is to improve reliability, a measure of reliability also has to be included in the trade-off context. Minutes of late time were therefore included in the choice scenarios. This would then also allow the estimation of the value of late time, which conveniently can then be used along with the value of recovery time to appraise measures to improve reliability through recovery time. We did not include any monetary terms in the SP exercise since it is sufficient that recovery time is valued in equivalent units of in-vehicle time to enable the extension of the railway industry's Generalised Journey Time (GJT) term to include a weighting of journey time in line with any premium valuation of recovery time. Based on the findings of the focus groups, recovery time was referred to as "contingency time" in the SP experiment offered to respondents.

An example of the SP experiment is provided in Fig. 1. It presented travellers with nine choices between two train service options characterised by different levels of scheduled and actual journey times, scheduled and actual contingency times and late times. Option A included contingency time whilst Option B had no contingency time for reasons of simplicity and offering a clear-cut trade-off. The reliability element was presented in what is now a fairly conventional manner of the journey times that might occur on five different days that was first introduced by Senna (1994). The actual journey times in combination with the planned contingency time imply an actual amount of contingency time and an element of late arrival time. These implicit figures were presented in the SP exercise to make the choice task clearer for respondents. For Option A, which contained the recovery time, the train journey times were chosen so that there were never any early arrivals.[5] The actual (unused) contingency time could be wait time between a

---

[5] This was to avoid adding a further dimension to an already challenging SP exercise given that valuing early station arrivals was not a requirement of the study. Where in reality recovery time manifests itself in early arrivals, it is usually at terminus stations although on longer distance journeys, such as on cross-country routes, there can be dwell times at larger stations along the route. On average, around 20% of trains in Britain arrive earlier than the publicly-advertised times but this is rarely by more than a minute or two since trains are generally held to their advertised departure times at all stations along the route.

pair of stations, as in the example of Fig. 1, or wait time at an intermediate station, and these were described in the introductory rubric and randomly allocated to respondents. An on-schedule journey would therefore imply an actual amount of contingency time equal to the scheduled amount. Option B, which has no planned contingency time, has a late time which is the actual minus scheduled journey time, and again no early arrivals were offered.[6]

The SP design needs to be based around the utility function that it is intended to estimate. This takes the form:

$$U_j = \alpha IVT_j + \beta R_j + \lambda L_j \qquad (1)$$

where for alternative j, IVT is the 'standard' in-vehicle journey time ('standard' means it excludes recovery time and late time), R represents the recovery time experienced and L denotes the late time. These are averages given that the choice context here covers 5 'daily' scenarios.

We adopted the 'boundary ray' approach to the design of the SP experiment (Fowkes 1991) on the grounds that this is both a feasible and attractive option when dealing with three attributes. This method aims to offer trade-offs across variables that are sensible in terms of the range of preferences that respondents might reasonably be expected to have. Setting α in Eq. 1 above to one, to operate in terms of the time multipliers we wish to estimate, the point of indifference, or boundary value, in the choice between options A and B is:

$$IVT_A + \mu R_A + \rho L_A = IVT_B + \mu R_B + \rho L_B \qquad (2)$$

where μ is the time value of recovery time and ρ is the time value of late time. We can therefore plot the relationship of indifference (boundary ray) between the value of recovery time (μ) and the value of late time (ρ). This is:

$$\mu = \frac{IVT_B - IVT_A}{R_A - R_B} + \frac{L_B - L_A}{R_A - R_B}\rho \qquad (3)$$

We specified option B to have no recovery time and hence to be the more unreliable option. The intercept is therefore $\frac{IVT_B - IVT_A}{R_A}$, with a slope of $\frac{L_B - L_A}{R_A}$. A respondent's choice indicates which side of the boundary ray they are located, and the design task is to select appropriate differences in time, late time and recovery time to offer a sensible range of choices.

The slope must here be positive. The intercept can be either positive or negative depending upon the sign of the in-vehicle time difference ($IVT_B - IVT_A$) which gives an element of flexibility in how the boundary rays cover the expected range of values. So how do we decide the differences in IVT? Given option B has more delay time which is relatively highly valued in terms of equivalent units of in-vehicle time, and given that we do not expect the premium attached to the greater recovery time in option A to be as large, then

---

[6] At the time of the survey delay-repay schemes compensate travellers for late arrivals of 30 min or more and so could have had a confounding effect on the interpretation of the SP options. The 9 SP scenarios offered to each person, which contained 90 levels of delay time in total, were therefore restricted to just two (out of 90) instances of 30 min delay and none greater.
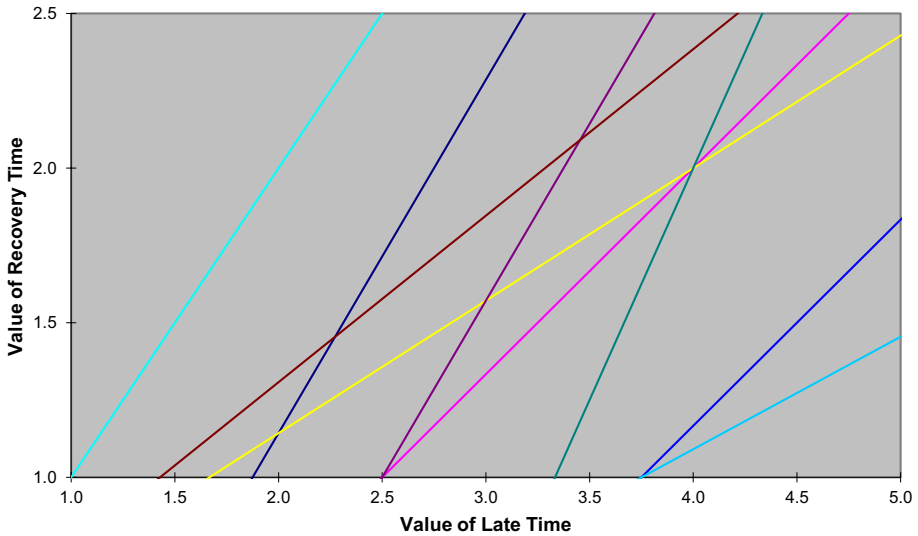
Fig. 2 Boundary ray map for 1 h design

we have made option B generally quicker so that sensible trade-offs are offered which yield useful information for modelling purposes.

We started with a design that was orthogonal in differences, with three levels of difference for each variable, and changed the attributes within the bounds of what we felt reasonable to obtain a sensible set of boundary rays. In the process, we departed from orthogonality but we ensured that the correlations between attribute differences were not large. An example of the boundary rays, relating to journeys of around an hour, is provided in Fig. 2.

At the time of the study, we were aware of other advanced SP design methods—e.g. D-efficient designs—which, in principle, yield more precise coefficient estimates (Bliemer and Rose 2011). Notwithstanding this, we used the boundary ray approach outlined above because it uses priors relating to relative valuations in determining the trade-offs to offer, and in the absence of previous work in the area we felt we could better 'guesstimate' relative valuations than coefficients. We also conducted simulation tests on our designs using synthetic choice data prior to implementation which indicated that they could accurately recover the relative values used in creating the test data. For future studies, however, a more efficient design could lead to the estimation of significant values with a lower sample size.

Separate SP designs were developed based around the actual journey times on the train routes selected to be surveyed. Such customisation is important for the realism required to obtain reliable responses given that the survey took the form of self-completion 'pen and paper' questionnaires. In total, six SP exercises were designed, based upon reported journeys of around 30 min, 45 min, 1 h, 1½ h, 2 h 15 min and 3 h. So, for example, the 2 h 15 min design catered for the Leeds/Wakefield to Kings Cross journey on the East Coast Mainline.

The scheduled recovery times were 5, 10 and 15 min except for the 30-min design where they could be only 5 or 10 min whereas 20 min was also permitted in the designs for journeys of 2 h 15 min and 3 h. While these levels may seem higher than what is typically observed in reality and accordingly people's perceptions (see Table 3), there are two reasons

to justify this selection. First, one of the objectives of this study is to inform decisions on how much recovery time to build in and whether adding more is desirable; given that recovery time will be readily appreciated to be under the control of train operators, then variations in it are credible provided that we limit it to what can be expected to be reasonable. Second, using too small changes in time is typically avoided in time valuation literature since respondents struggle to make meaningful choices (this is discussed in more detail below as it also applies to other attributes). Furthermore, one of the purposes of the focus groups was to explore recovery time levels of these magnitudes and they were deemed to be realistic.

The levels of late time ranged from zero to 30 min, and also included 5, 10, 15 and 20 min. Where recovery time existed, late time took the level zero in four or five instances in all but one scenario where three of the five late times were zero. When there was no recovery time, the pattern was two of the five late times were zero. The non-zero late times tend to be much larger than typical or average, and as a consequence the mean lateness in the SP exercises of between zero and 9 min where there is recovery time and between 3 and 12 min where there is no recovery time tend to exceed the mean lateness on the routes we were to survey which ranged from 1.8 min on the shorter distance flows covered by South Western Trains through to 3.7 min on the longer distance journeys served by Cross Country. The reasoning behind the late time levels adopted is that restricting them to the few minutes of actual circumstances would run the risk that the variations would have been ignored. Indeed, offering somewhat larger amounts of late time than are routinely experienced is customary in SP studies of travel time variability, such as in the UK (ARUP et al. 2015) and Dutch (Significance et al. 2012) national value of time studies.[7] Moreover, the levels of late time (as well as those for recovery time) and distributions used in the SP exercises had been explored in the focus groups and no adverse feedback was received, whilst we incorporated the safeguard of asking respondents how realistic they found the levels offered whereupon we can test the impact on the estimated parameters of perceived unrealism.

SP data collection and summary statistics

The SP surveys were conducted on train, partly to ensure passengers were thinking of an actual journey when completing the SP questionnaires and partly because this is a very cost-effective means of achieving large samples. Table 1 indicated that the on-train sample corresponds closely with the NTS in terms of gender, age group, employment status and occupation. Whilst "The online and on-train surveys" section pointed out differences in the journey purpose distributions between the two samples, we account for this in our analysis of the SP data through journey purpose segmentations.

There was a 66%:34% split between those choosing option A, which contained recovery time, and Option B, which contained none, reflecting an overall favourable view of the inclusion of recovery time in timetables. When asked about the realism of the SP exercise, 5.5% reported the journey times to be fairly unrealistic with 1.5% regarding them to be very unrealistic. The corresponding figures for late time were 10.0% and 1.6% and for wait time were 19.0% and 3.4%. It is not surprising that wait time is regarded to be most unrealistic since it is likely to be the variable that respondents are least familiar with. Nonetheless, the attribute levels were largely regarded to be realistic and when perceived unrealism was allowed to impact on the relevant parameter estimates no significant effects were

---

[7] It is also customary in value of time studies that time variations of less than 5 or 3 min are not offered even though values of time are used to evaluate project time savings very much less than this.

obtained. Respondents did not appear to have undue problems with the SP exercise, with around a third finding it to be very easy and only 3% reporting it to be very difficult.

Analysis of stated preference data

Often, SP studies resort directly to model estimation to analyse individuals' choices. Such an approach, while efficient, risks missing part of the story hidden in the dataset. We believe this is especially the case with this study, as it is the first analysis of recovery time from the passengers' perspective. Thus, we first report a preliminary analysis of the dataset (5.1), followed by a description of the choice modelling methodology (5.2) and the main results (5.3).

Preliminary analysis

Non-trading behaviour

There was a high degree of *non-trading* in the sample, with 454 respondents (47%) always choosing the same option across all 9 choice tasks. This is made up of 37% who always chose option A, which includes the recovery time, and 10% who always chose the less reliable option B. However, while the term "non-trader" is widely used in the choice modelling literature, technically we cannot say whether these people did trade-off between options or not. What we mean is that it is possible that these respondents did trade-off, but concluded every time that option A or B was preferable for them based on their intrinsic valuation of the attributes involved. It is also known that non-trading might also be related to inertia effects (Hess et al. 2010); in this respect, while none of the options represents the exact current trip of passengers, it is possible that those who do (do not) currently perceive some recovery time are inclined to the option with (without) recovery time. At the other extreme, non-trading behaviour can also be linked to lack of engagement or experienced difficulty with the survey, leading to the same choice in every task (although the latter is unlikely considering respondents' feedback on the survey exercise).

One way of finding out is to check whether non-trading behaviour is related to intrinsic attitudes and preferences. Looking at these respondents in more detail, which we do in Table 5, their choices are generally consistent with their personal view on the use of recovery time by train operating companies. Those who approve of recovery time are more likely to always choose A (45%) in comparison with those with different views (3%, 8%

Table 5 Personal views and non-trading behaviour

	Personal views regarding the use of recovery time				Total
	Approve	Neither	Disapprove	Cannot say	
Trading behaviour					
Trade	51%	56%	55%	65%	518
Always A	45%	35%	8%	26%	356
Always B	4%	9%	37%	9%	98
Total	550	256	123	43	972

and 26% respectively) whilst those who disapprove are more likely to always choose B (37%) compared to those with other views (4%, 9% and 9% respectively).

At the same time, more than 50% of respondents within each 'personal view' category chose A or B at least once. Overall, the information provided by 'non-traders' is justifiable on the grounds of being correlated with their personal views and we cannot discard it. Nevertheless, having half of the sample always choosing the same option can be a problem for the estimation of the model parameters, and this will be taken into account at the modelling stage.

Choice patterns by key characteristics

It is well known that the rail travel market is made up of a number of different segments with often distinctly different preferences. Travel demand studies generally distinguish between commuters, business travellers and leisure travellers and we do that here. Table 6 provides, separately by journey purpose, further insights into the pattern of SP responses. The statistics shown help to explain how preferences towards recovery time vary depending on journey purpose.

A number of key messages emerge from Table 6, starting with some evidence of positive preference towards recovery time, judging by the proportions of choice and personal views. In terms of choice, all three categories of journey purpose have a sizeable majority of respondents who prefer option A (with recovery time). This ranges from 60% for commute to 75% for leisure. In terms of personal views, over 80% of all respondents either approve (57%) or neutral (30%) about the use of recovery time on rail.

Responses between different journey purposes in Table 6 come through strongly. Leisure travellers chose the option with recovery time 75% of the time and 60% of them approved of its use. In contrast, among commuters "only" 60% chose option A and 53% approved of its use. In between these lie business users with 63% choosing option A and 58% approving of recovery time.

Such differences reflect the characteristics of these individual travel markets and are a positive validation of the survey results. For example, leisure trips are typically one-off trips and sensitive to punctuality (going to the theatre, meeting friends for dinner, catching

Table 6 Choices and key aspects of preferences by journey purpose

| | | Journey purpose | | | |
		Commute	Business	Leisure	Total
Choice					
Option A—recovery time	Obs	1838 (61%)	1854 (63%)	2035 (75%)	5727 (66%)
Option B	Obs	1169 (39%)	1108 (37%)	669 (25%)	2946 (34%)
Personal views					
Approve	Indivs.	174 (52%)	193 (58%)	183 (60%)	550 (57%)
Neither/cannot say	Indivs.	103 (30%)	95 (29%)	101 (33%)	299 (30%)
Disapprove	Indivs.	60 (18%)	44 (13%)	19 (6%)	123 (13%)
Awareness					
Not aware	Indivs.	176 (52%)	192 (58%)	184 (61%)	552 (57%)
Aware	Indivs.	161 (48%)	140 (42%)	119 (39%)	420 (43%)
Total	Indivs.	337	332	303	972
	Obs	3007	2962	2704	8673

a flight etc.). As such, any delay to a leisure journey will have a strong disutility for that passenger and it is logical that they will have a stronger preference for recovery time.

Commuters travel much more frequently. Consequently, they are likely to experience unreliability on a regular basis and may either have contingency plans for such events or build in their own recovery time by catching an earlier train than they need to. For some commuters (note not the majority) there may be a preference for shorter journeys above more reliable journeys, e.g. saving 15 min 4 days of the week and risk being late 1 day is preferred to no travel savings and guaranteed punctuality. This preference may differ by the type of job (nurse vs an office worker), flexibility of the work environment and the individual. Overall however, the majority of commuters are still supportive of recovery time, albeit less so than leisure travellers.

A business trip can share the features of commuting and leisure trips. In terms of frequency it is more likely to resemble a leisure trip, but it also shares with a commuting trip the consequences of delay, e.g. being late to a meeting, which may be perceived as more palatable than the delay consequences in a leisure trip, e.g. missing the start of a concert or a holiday trip. It is therefore not surprising that the level of approval of recovery time for business travellers lies in between those for leisure and commuters.

Valuation methodology

Choice models are used to analyse the data and infer valuations. First, we describe the base functional form of the Multinomial Logit (MNL) model used, which is followed by two sets of extensions that allows us to account for observed heterogeneity across passengers and the impact of time variability. All preferred models reported in the paper include observed heterogeneity.

Base model

Individuals are assumed to make a choice between the two options in order to maximise their utility. Each of the two travel options j has an associated utility U_j that is defined in terms of the attributes presented in the SP experiment. The utility function of the base MNL model is specified as follows:

$$U_j = \beta_{ivt}E\left(IVT_j\right) + \beta_{late}E(LATE_j) + \beta_{rrt}E(RRT_j)$$ (4)

Since the two options were defined for 5 typical trips over a week, the expected value of each attribute across the 5 trips is used. $E(IVT_j)$ is the average standard In-Vehicle Time for alternative j, $E(LATE_j)$ is the average late time for alternative j, and $E(RRT_j)$ is the average residual recovery time for alternative j. Hence, the RRT is that part of the built-in recovery time that is left 'unused' where it was not replacing any late minutes. In other words, RRT is the additional cost (in time units) that passengers incur in return for reduced late time. The β_{ivt}, β_{late} and β_{rrt} are parameters to be estimated, each indicating the marginal utility of an additional minute for the three types of travel time minutes described above.

Three key valuation measures can be inferred from the observed choices in this model:

- The value of mean residual recovery time (VMRRT), giving the value of 1 min of mean residual recovery time in relation to 1 min of mean standard in-vehicle time:

$$VMRRT = \frac{\beta_{rrt}}{\beta_{ivt}}$$ (5)

In other words, the VMRRT is a multiplier on IVT that measures the premium, as perceived by travellers, of a minute of additional time inside the train due to the introduction of recovery time in the timetable.

- The value of mean late time (VML), which is a conventional measure of the value of reliability, gives the value of a minute of mean late time in units of mean IVT:

$$VML = \frac{\beta_{late}}{\beta_{ivt}} \qquad (6)$$

  This is known in the railway literature as the late time—or lateness—multiplier (Wardman and Batley 2014).

- The ratio VML/VMRRT gives the weight of 1 min of mean late time relative to 1 min of mean residual recovery time. This ratio can be interpreted as a type of late time multiplier. Because recovery time is aimed at reducing late time, this is the key measure that informs on how many minutes of mean residual recovery time people are willing to accept in order to reduce mean late time by 1 min.

Due to our interest in valuation estimates, we transform the model into 'valuation space', analogous to what is known as Willingness-to-Pay space in the literature on valuation of travel time (Train and Weeks 2005). Given that the VML and VMRRT have a common denominator (β_{ivt}), the model can be translated into a valuation space where we readily obtain estimates of the VMRRT and VML in minutes of IVT:

$$U_j = \beta_{ivt} * \left[E\left(IVT_j\right) + VML * E(LATE_j) + VMRRT * E\left(RRT_j\right) \right] \qquad (7)$$

This model is equivalent to that of Eq. (4), but it provides direct estimates of VMRRT and VML. The coefficient on IVT (β_{ivt}) becomes effectively a scale parameter in this model. This has the advantage of facilitating the modelling of preference heterogeneity, which can now be linked directly to the valuation estimates (VML and VMRRT). In other words, any preference heterogeneity in the relative valuation of RRT and IVT can be reflected directly through VMRRT instead of through two separate coefficients as in more traditional models (Train and Weeks 2005).

Accounting for heterogeneity in travellers' preferences

Valuations of recovery time (VMRRT) and late time (VML) are likely to be heterogeneous across respondents. An extensive search for a model specification that dealt with observed and unobserved heterogeneity was conducted, including many variations of MNL, Latent Class (LC), and Mixed Logit (MMNL) models (McFadden and Train 2000). Four different extensions of the MNL model, all of which account for observed heterogeneity, are reported in "Results" section below.

Before the selected models are presented, we briefly summarise the specification search and why the other options (namely LC and MMNL variations) were discarded. Passengers' preference heterogeneity can be accounted for in the model through observable variables, such as journey purpose and length, but also in the random error term structure since part of the heterogeneity will not be observable.

The Mixed Logit (MMNL) model is typically useful to estimate a distribution of values. We tried and discarded this option because we could not successfully estimate a MMNL model that resulted in significant set of segmented values (e.g. by trip purpose). We feel that

this is partly because the values of recovery time and mean late time are multipliers of time, and we would not expect variations in multipliers to be as large as for the more typically estimated monetary values—making it more difficult to simultaneously pick up random and observed heterogeneity. This is in line with the literature, including the most recent studies that estimate late time multipliers (e.g. Hess et al. 2017). We also suspect that the high degree of non-trading in the data limited the possibilities of recovering distributions of values jointly with observed heterogeneity through a MMNL.

The latent class (LC) model was technically a better option than the MMNL as a behavioural model, since instead of a distribution of values, the LC estimates a discrete set of different values for each of the latent classes of respondents, where both parameters and class allocation could be associated with individuals' characteristics. Nevertheless, the LC was also discarded for the same reason, namely the failure to provide a significant set of segmented values within a LC structure. Again, we believe this could be linked to the high-degree of non-trading in the data, which limits the possibilities for unpacking all sources of heterogeneity.

The estimated MNL models relate all observable heterogeneity to the valuation measures VMRRT and VML in a way that sample average values can be derived for all segments of interest (e.g. commute, business and leisure). Additionally, it is possible to partially control for non trading behaviour. The variables tested include awareness of recovery time, gender, age, journey duration, journey purpose, the need for interchanges and personal views on approval/disapproval of recovery time.

Discrete and continuous multipliers were added to the VMRRT and VML parameters to discern different valuations for different segments. The use of multipliers, instead of additive interactions, simply facilitates direct calculation of valuation estimates for different segments. For discrete variables (e.g. purpose), multipliers on VMRRT and VML were estimated for all categories (e.g. commute and leisure) except for one, the base (e.g. business), for which its multiplier was set to 1. Using travel purpose multipliers on the VML as an example, we specified:

$$VML * \left(\delta_{business} + \theta_{VML_{com}} \delta_{com} + \theta_{VML_{lei}} \delta_{lei} \right) \tag{8}$$

where δ_x are the dummy variables for the different x categories of the variable, and θ_{V_x} are the valuation multipliers for valuation measure v and category x. In this example, business ($\delta_{business}$) is the base category, and $\theta_{VML_{com}}$ and $\theta_{VML_{lei}}$ the multipliers for categories commuting and leisure on the VML.

The only continuous variable was journey duration and its multiplier was entered as an elasticity specification, commonly used in valuation studies (Mackie et al. 2003; ARUP et al. 2015). Taking VMRRT as an example:

$$VMRRT * \left(\frac{D}{\bar{D}} \right)^{\lambda_{D\_VMRRT}} \tag{9}$$

where λ_{D_VMRRT} is the journey duration elasticity of VMRRT to be directly estimated. The denominator $\bar{D}$ in the multiplier ensures that the readily estimated VMRRT relates to a train journey of $\bar{D}$ minutes. This $\bar{D}$ value was set to 45 min for commuters and to 90 min for leisure and business travellers.[8] Since separate elasticities were estimated by purpose, $\lambda_{D\_V\_x}$ would represent the distance elasticity on valuation v for purpose x.

---

[8] The choice of denominator does not affect the valuation estimates.

Last but not least, an elegant solution was adopted to deal with the high degree of non-trading in our sample. As we previously argued, discarding their preferences was not an option because non-trading behaviour was to some extent related to personal views on the use of recovery time (see Table 5). On the other hand, including multipliers for the non-traders was also not an option because their valuations would not be correctly identified. We needed a mechanism that somehow allowed us to partially control for their extreme preferences, but without fully removing them from the mean valuation estimates. Since the variable 'personal views on the use of recovery time' and non-trading behaviour were partially correlated, the solution adopted was to include multipliers on the valuation measures (e.g. VML) for the "approvers" and "disapprovers" categories (neutral views was set as the base category). These multipliers—modelled in the fashion of Eq. (8)—capture any excess (above or below) on valuations that is strictly related to personal views. As we shall see, with this approach, valuation measures fall somewhat in-between those of the model without these additional multipliers and a model which excludes non-traders. Results for these two additional model variations are included (models 2 and 3) and discussed in the results section.

Accounting for the variability of time

So far, the models only use the expected values of the time distributions but disregard the variability of outcomes. However, the wider the spread of potential travel time outcomes, the higher the risk and uncertainty. A final extension of the base model (model 4) deals with the variability in time.

In this extension, we move away from the assumption in previous models that each of the five outcomes (see Fig. 1) has an equal weight. This version of the model allows for different weights for the five travel time outcomes using a constant absolute risk aversion (CARA) specification (Liu and Polak 2007; Hess et al. 2017):

$$U_j = \beta_{ivt} * \left[\sum_s \frac{1 - e^{-\alpha V_{j,s}}}{\alpha} \frac{1}{5} \right] \tag{10}$$

where

$$V_{j,s} = IVT_{j,s} + VML * LATE_{j,s} + VMRRT * RRT_{j,s} \tag{11}$$

where s indicates the travel outcome ($s=1,\ldots,5$) and j indicates the travel alternative. The new parameter α picks up the degree of risk aversion/seeking behaviour: a positive (negative) α indicates risk aversion (seeking) attitude. Behaviour is risk neutral when α approaches zero, thus approximating the base model from Eq. 7.

A second extension using a mean-standard deviation model was tested, adding a coefficient on the standard deviation of late time as an additional term.[9] However, the high correlation between mean lateness and the standard deviation (around 95%) did not allow us to separate the two impacts and the CARA approach was preferable. This was also the

---

[9] We would though point out that only between 1989 and 1997 PDFH contained recommendations relating to the standard deviation of late time and these were dropped in version 3 as the evidence based was deemed to be insufficiently robust.

approach chosen to account for time variability in the latest UK national study valuing in-vehicle travel time and late time (Batley et al. 2017; Hess et al. 2017).

Results

The estimation results are presented in Table 7 below. Following an extensive model specification search, four models are reported. A base model (model 1) is reported first. Model 1 accounts for part of the observed heterogeneity in the data but critically does nothing in relation to the presence of non-traders. The same model is also estimated including a treatment of the non-traders' choices (model 2) via the multipliers of approval views (θ_disapprove and θ_approve) and excluding the non-traders (model 3) for comparison. Model 4 builds upon model 2 to take account of travel time variability through a CARA specification.

As expected from our preliminary analysis of non-trading behaviour, the relatively poor explanatory power of model 1 is arguably related to the lack of any control for the non-traders' responses. Model 2 significantly outperforms model 1 thanks to the valuation multipliers on passengers' approval/disapproval of recovery time. An overview of the results of models 1, 2 and 3 show that, as expected, model 2 (with controls) derives valuations that are in-between those from the model which excludes non-traders (model 3) and the model that includes them but do not control in any way (model 1). Finally, model 4 produces outcomes that are not too far from those of model 2 but it is a more flexible specification which allows for differential weights for each of the five travel time outcomes. The positive *α_CARA* estimate suggests the presence of risk averse behaviour in the sample. Moving forward, considering the statistical superiority of models 2 and 4 and the greater flexibility of model 4, we focus on the outcomes from model 4.

The value of mean late time (VML) indicates that business travellers are willing to accept 5.80 min of IVT if this reduces mean late time by 1 min. For commuters and leisure travellers, the VML is equal to 5.19 and 5.78 respectively but both θ_*VML_leisure* and θ_*VML_commute* multipliers are not significantly different from 1 and thus the difference across purposes is not statistically significant. These estimates are higher than current recommendations of between 2.3 and 3.9 min (these vary by segments) in UK official guidelines for appraisal and forecasting but not far from existing evidence in the literature (Wardman and Batley 2014). The recommended late time multiplier in the UK is higher than 3 in cases of rail links to airports or long distance journeys. Moreover, Abrantes and Wardman (2011), in a meta-analysis of valuation of time studies in the UK, found a mean multiplier of 6.35 across 15 observations.

The value of mean residual recovery time (VMRRT) is estimated at 1.88 min of IVT for business travellers, 1.46 for commuters, and at 0.79 for leisure travellers; however, these differences across purposes are not significant. Taking the value of 1.88 estimated on business travellers, this indicates that 1 min of recovery time is perceived like 1.88 min of IVT. While these estimates are only significantly different from 1 at the 90% level of confidence, this is only the case for the underlying base category: other multipliers (e.g. awareness or journey duration) modify these estimates for specific segments.

All other multipliers are statistically significant and can be interpreted as follows. Awareness of recovery time (θ_*aware*) is associated with higher VMRRT, which means that people who are aware of recovery time associate it with a more negative utility than IVT. We find this credible since aware passengers will have more experience of the frustration involved in prolonged station stops, stopping outside stations and slowed-down

Table 7 Estimation results

Model	1		2		3		4	
Model description	Base		Non-traders treatment		Excl. Non-traders		CARA specification	
	Est.	t-ratio[a]	Est.	t-ratio[a]	Est.	t-ratio[a]	Est.	t-ratio[a]
β_ivt	−0.04	−1.31	−0.04	−5.97	−0.08	−9.27	−0.07	−5.48
		t-ratio (1)		t-ratio (1)		t-ratio (1)		t-ratio (1)
Base valuations and purpose multipliers								
VMRRT (base = business)	1.05	0.04	1.58	1.29	2.08	4.54	1.88	1.95
VML (base = business)	5.53	2.77	4.83	4.98	4.33	8.36	5.80	5.47
θ_VMRRT_leisure	0.82	−0.12	0.44	−1.19	0.99	−0.06	0.42	−1.50
θ_VML_leisure	1.54	1.71	1.09	0.47	1.06	0.49	1.00	−0.01
θ_VMRRT_commute (for a 45 min. trip)	1.05	0.06	0.99	−0.03	0.76	−1.77	0.79	−0.95
θ_VML_commute	0.99	−0.02	1.09	0.6	0.73	−2.01	0.89	−0.68
Other multipliers on VMRRT								
θ_aware	2.39	0.40	1.63	1.66	1.15	1.61	1.57	1.93
θ_interchange	0.55	−0.63	0.56	−2.2	0.87	−1.30	0.58	−2.34
Other multipliers on VML								
θ_approve			1.55	3.91			1.65	3.86
θ_disapprove			−0.25	−7.37			−0.11	−8.87
		t-ratio (0)		t-ratio (0)		t-ratio (0)		t-ratio (0)
Journey length elasticities								
λ_D_VMRRT_commute	0.30	1.11	0.49	2.47	0.45	2.01	0.60	2.96
Time variability parameters								
α_CARA							0.003	4.79
Final log-likelihood	−6011.6		−5144.36		−3046.8		−5144.12	
Adj ρ2	0.08		0.14		0.04		0.14	
Observations	8673		8673		4587		8673	
Respondents	972		972		518		972	

[a]t-ratio (0) is the standard t-ratio. t-ratio (1) is used for VMRRT, VML and other multipliers, to determine whether they are significantly different from 1

running elsewhere. Passengers who need to interchange perceive recovery time more positively relative to IVT (as indicated by θ_*interchange*) probably because they are more likely to benefit from it than other passengers. On the other hand, the VMRRT varies across different journey durations for commuters but not for other purposes (thus, only the journey time elasticity λ_*D_VMRRT_commute* is included in the final models). The longer the journey, the less beneficial recovery time is for commuters; for shorter journeys, recovery time is perceived practically the same as IVT (I.e. VMRRT close to 1; see Table 8). This is a reasonable result as long commutes are burdensome. People with long commutes will still want reliable journeys, but not so much at the expense of additional IVT (i.e. they may prefer solutions other than increased recovery time).

Finally, it can also be observed that VMRRT in model 4 (and in model 2) fall in-between the base VMRRT from the 'no-treatment' base model (model 1) and the 'excluding non-traders' model 3. When non-traders are excluded, VMRRT is roughly equal to 2.1 and significantly different from 1; when non-traders are included but not controlled for VMRRT is equal to 1 (this was expected since most non-trading was in favour of the recovery time option). This confirms that our 'treatment' model provides somewhat of a compromise solution between the extreme options of excluding non-traders and not doing anything about it. Nonetheless, more importantly the overall issue with non-trading detected in this (first) study forces us to strike a word of caution when interpreting the results and provides a valuable lesson for the further studies that examine recovery time preferences.

The multipliers on personal views (preferred model) are highly significant and contribute to a large improvement in model fit for models 2 and 4. As expected, the approval (disapproval) multipliers are capturing the much higher (lower) VML of people who always chose the option with recovery time (without recovery time). Unfortunately it is not possible to know the extent to which these multipliers remove the impact of non-trading, but they prove very helpful in allowing us to understand their influence. Also, another advantage of their introduction in the model is revealed by looking at the effects of awareness and interchange across the first three models. There seems to be some confounding between personal views and whether people are aware or have to interchange, and only the inclusion of the personal views multipliers allows all these effects to be disentangled in the model, increasing the precision of the estimates on the awareness and interchange multipliers.

Other multipliers on VMRRT and VML (e.g. distance, age or gender) were also tested but found to be not significant and removed from the final specification. Table 8 summarises the main valuations by key segments, including the VML/VMRRT ratios, using estimates from model 4.

For each travel purpose we estimate VMRRT, VML and their ratio for the following four categories: unaware (base), aware, awareness weighted average and interchange. Three of these categories follow straightforwardly from the model estimates, whereas the 'awareness weighted average' one provides a weighted average of the values of aware and unaware passengers, using the weights from the sample (reported in Table 6).

The discussion of these results focuses on the aware/unaware weighted average category as being an interesting representative case of the overall sample. Table 8 shows how leisure travellers have the lowest VMRRT, equal to 1, meaning that recovery time is deemed to be just like IVT and no extra penalty is associated to it. Business travellers have a VMRRT of 2.3, while commuters' values vary between 1.2 for very short journeys and 2.9 for very long ones. These values are the first to be estimated on this aspect of rail travel and seem reasonable considering the different nature of each travel purpose and the insights from the wider study and focus groups (prior to the SP experiment).

Table 8 Valuations by segment (model 4)

Journey duration (mins.)[a]	Unaware			Aware			Weighted average awareness			Interchange		
	Commute	Leisure	Business	Commute	Leisure	Business	Commute	Leisure	Business	Commute	Leisure	Business
VMRRT (min/min)												
20	0.91			1.42			1.2			0.53		
30	1.16			1.82			1.5			0.68		
(base) 45	1.48	0.79	1.88	2.32	1.24	2.94	1.9	1.0	2.3	0.86	0.46	1.10
60	1.76			2.76			2.2			1.03		
90	2.25			3.53			2.9			1.31		
VML (min/min)												
	5.19	5.78	5.80	5.19	5.78	5.80	5.19	5.78	5.80	5.19	5.78	5.80
VML/VMRRT (lateness-recovery ratio) (min/min)												
20	5.7			3.7			4.5			9.8		
30	4.5			2.9			3.5			7.7		
(base) 45	3.5	7.3	3.1	2.2	4.7	2.0	2.8	6.0	2.5	6.0	12.5	5.3
60	2.9			1.9			2.3			5.0		
90	2.3			1.5			1.8			3.9		

[a]Values vary by journey length only for commuters. Unique values apply to other purposes

We now turn to the ratio VML/VMRRT, which for simplicity we will refer to it as Lateness-Recovery (LR) ratio. This metric is important because it can be interpreted as a pseudo-lateness multiplier that is more relevant for policies that add or reduce the amount of recovery time in a corridor. The LR ratio contrasts 1 min of late time, instead of with additional IVT, with additional residual recovery time. Thus, it compares late time with residual recovery time, which is the relevant information for the appraisal of such policies. The estimated LR ratios are 2.5 for business travellers, 6 for leisure travellers, and between 1.8 and 4.5 for commuters depending on trip duration. These values indicate the minutes of mean residual recovery time that passengers are willing to accept in return for a reduction of 1 min of mean late time. The higher the ratio, the greater the case for recovery time, and vice versa. Our results confirm that recovery time can be especially beneficial for leisure travellers, given the sensitivity of leisure plans to potential delays. On the other hand, recovery time will be least beneficial for commuters travelling over 1 h, as they would not be in a position to accept any extra travel time and would probably make their own adjustments to account for potential delays anyway. These values can in principle be used for appraisal. But, as this is the first study into the issue, it goes without saying that supplementary evidence will be required. This is even more so due to the high presence of non-trading. Further research should be conducted to consolidate the novel first set of values estimated in this study and should focus on reducing the high non-trading behaviour with an improved stated preference design. The outcomes from this first study can inform a better selection of trade-offs in future experiment designs.

One drawback of the results is the estimation of VMRRT lower than 1. This is only the case for the values from interchangers (other isolated cases can be observed in Table 8 but these are not significantly different from 1). It may be argued that it does not make sense that 1 min of travel time (in the form of recovery time) can be perceived as less penalising than 1 min of IVT. In that case, any VMRRT for use in appraisal should never be below 1. However, it might also be argued that recovery time is associated with an additional benefit (increased reliability) which is not assumed for IVT.

Comparison with the wider valuation literature

Since the estimates of VMRRT are the first evidence of its kind, they do not have an available direct comparator in the literature. The only and closest comparator can be found in the latest value of time study in the UK (ARUP et al. 2015), which provided values of 'slowed down' time for bus users. The multipliers for 'slowed down' time were found to be between 1.4 and 1.6, not far from our average estimates for recovery time.

When comparing this study with other studies of late time valuation in the railway literature (see Wardman and Batley 2014), our (high) estimates of late time multipliers are also not directly comparable as such for one reason. In contrast with previous SP experiments, ours is unique because it is the only one where the responsibility of delays is being explicitly held by the rail operator, at least to an extent. In other studies, the value of reliability was derived by asking people to pay higher fares or incur alternative longer journey times in return for reduced lateness. This might reveal a very different set of preferences because delays are not the responsibility of rail users and they may be reluctant to pay to reduce them. In this study, however, the way to reduce delays is through the operator's decision of including recovery time. This of course means a longer journey to the passenger, but also to the operator who might expect a demand reduction as a consequence. In some way, recovery time can be seen as an act of responsibility on the part of the operator. This might

be a reason why often respondents have chosen rail trips with recovery time and have been so supportive about it during the survey, hence revealing atypically high values of late time relative to both in-vehicle time and residual recovery time.

Illustrative use of values in appraisal

In this section we discuss how the valuation estimates provided could be used for appraisal. First, as set out in the context, the motivation of this research was to analyse whether perceived recovery time carries a premium relative to standard in-vehicle time. If so, this premium should be applied to changes in recovery time and, if greater than 1, would reduce the reliability benefits of policies that increases recovery time.

Perceived recovery time—also referred to as "residual recovery time" (RRT)—has been approximated in the study as "waiting time on board between or outside stations". The results show that the RRT carries a premium in some circumstances. RRT is valued between 1 and 3.5 times IVT, depending on journey purpose, distance, interchanges and awareness. However, it is not so clear that these premiums would translate into reduced reliability benefits in appraisal. The reason is that we have also found late time to carry a higher premium than the standard late time multipliers used in the UK (ATOC 2013), which are in a range of 2.3 to 3.9 (with the exception of trips to airports, where a multiplier of 6 is used). Altogether, we find that a minute of late time is valued at almost 3 times a minute of RRT for commuters and business travellers, and at 6 times for leisure travellers. This correspondence is not far from current practice, i.e. using existing late time multipliers on IVT without attaching a premium to perceived recovery time.

Consequently, we would not recommend the use of the estimated premiums for perceived recovery time in isolation under the current appraisal standards (where they would be combined with separately estimated late time multipliers). The trade-off between late time and recovery time is as important as the trade-off between IVT and recovery time, and they should be considered simultaneously. More precisely, from our study we observe that premiums for perceived recovery time (VMRRT) higher than 1 are associated with higher late time multipliers (VML). The VML/VMRRT ratios estimated show the relative values of late time and recovery time in our study. In practice, thus, some adjustment is likely to be necessary to align the relative values of IVT, recovery and late time.

Another implementation problem is that, in practice, it is not possible to separate "used RT" from the perceived "residual RT", as this would depend on the distribution of delays for each corridor. Therefore, a pragmatic solution is to assume that all RT will be perceived by passengers, which is a realistic assumption. It is realistic because a route with average lateness of 2 min typically does not have every train delayed by 2 min, but instead some trains are delayed longer while others are on time. Any RT introduced is likely to be perceived. Under this assumption, we seek to apply a premium to any additional RT.

Let us consider a policy that adds recovery time (ΔRT) and consequently changes (reduces) average minutes of late time (ΔAML). While $\Delta AML < 0$ is a benefit to passengers, $\Delta RT > 0$ is a cost. In the appraisal of such a policy, current practice uses IVT valuation for the additional RT and late time multipliers for the savings in AML. Our study provides new valuation estimates that can enhance the appraisal. As discussed earlier, if the VMRTT are going to be used in conjunction with external late time multipliers, the valuation premium applied to RT should be adjusted based on the values of the Lateness-Recovery (LR) ratio (i.e. VML/VMRRT). To infer the premium or multiplier for RT (namely,

RTM = recovery time multiplier), we can contrast the late time multiplier in use from the appraisal guidelines (VML$_{guidelines}$) with the recommended Lateness-Recovery ratio (LR) from this research. RTM can be calculated as follows:

$$RTM = Max\left(\frac{VML_{guidelines}}{LR}, 1\right) \tag{12}$$

The RTM is set to be at least equal to 1. Values lower than 1 are difficult to justify, since they imply that RT is less penalising than standard IVT and this is unrealistic. This formula is provided because it is not possible to offer direct estimates that match the segmentation of late time multipliers used in in UK appraisal guidelines (see PDFH; ATOC 2013) and because it has the advantage of being applicable to any other established late time multipliers (e.g. future updated guidelines). As an illustration, if we take an external VML from guidelines equal to x (e.g. 3), then the RTM will only be higher than 1 if, for a given segment, the LR (as reported in Table 8) is lower than x (e.g. 3).

Evaluation of an actual timetable change

We have obtained data for a scheme in Great Britain which introduced recovery time (RT) into railway timetables and subsequently identified the impact on reliability in terms of mean late time (AML). Analysis was carried out for a number of service groups (SGs) (train services sharing a common route section). SGs 2, 3 and 5 are suburban/local in nature, SG4 is outer-suburban (including some significant intermediate centres generating commuting and business traffic) whilst SG1 is a longer-distance route with more leisure traffic. To illustrate the different estimates across purposes, we calculate the changes in generalized journey times (GJTs) for each purpose separately. The applicable LR ratios (VML/VRRT) for leisure and business travel are equal 6 and 2.5 respectively, based on the estimates provided in Table 8 above, whereas for commute travel the LR ratios vary by journey length.[10] For simplicity, and to avoid releasing anonymous details of the schemes analysed, we use the prevailing average late time multiplier (VML$_{guidelines}$) from PDFH (ATOC 2013) at the time of this research, equal to 3. A late time multiplier of 3 is in line with some of the U.K. evidence reviewed by Wardman and Batley (2014) and is widely used internationally (OECD/International Transport Forum 2014).

With this information, we can calculate the RTM (Eq. 12). For example, for business travel, the RT should be valued at 1.2 times the IVT (RMT = 3/2.5 = 1.2). With the RTM and VML, we can evaluate policies that increases RT in exchange for reduction in the Average Minutes of Lateness (AML), attaching a premium to recovery time if applicable. We obtain the difference in Generalized Journey Time (GJT) between the Base situation and the Do-Something (characterized by longer IVT—due to added RT—and lower AML) for the five service groups. This exercise is summarised in Table 9 below.

We first observe that adding RT did result in significant reductions in AML for the five cases as a result of the scheme. Using VML = 3 and our estimated RTMs, Table 9 show significant reductions in weighted GJT in all cases, albeit less so than with the current approach (which implicitly assumes RTM = 1, i.e. recovery time is valued just like in-vehicle time in all cases). As it stands, aligning our LR with guidelines late time multipliers of

---

[10] The LR ratios for commute are applied based on journey length (as per Table 8) using the base IVT for each service from Table 9.

Table 9 Examples of actual timetable changes in the UK: calculation of changes in GJT using the estimated recovery time multipliers (RTM)

Service group	Base case		Do-Something (with RT)		RTM			Net changes		ΔWeighted GJT[a]		
	IVT	AML	IVT+RT	AML	Comm	Leis	EB	ΔRT	ΔAML	Comm	Leis	EB
1	60.7	7.8	64.4	2.3	1.3	1	1.2	3.7	−5.5	−11.7	−12.8	−12
2	38.5	3.4	42.3	0.8	1.09	1	1.2	3.8	−2.6	−3.7	−4.0	−3.2
3	32.1	4.6	34.4	1.9	1	1	1.2	2.3	−2.7	−5.8	−5.8	−5.3
4	58.2	3.6	59.9	2	1.3	1	1.2	1.7	−1.6	−2.6	−3.1	−2.8
5	22.6	4	23.7	1	1	1	1.2	1.1	−3	−7.9	−7.9	−7.7

[a] To calculate ΔWeighted GJT, ΔRT is multiplied by the RTM and ΔAML by the VML (equal to 3)

Comm, commuting; Leis, leisure; EB, business

3, the recovery time premium is greater than 1 for business travel and medium to long commutes. In those cases, the GJT reductions relevant for appraisal would differ. The impacts will be larger if higher late time multipliers were considered. To calculate the monetary benefits of these policies, the change in weighted GJT would need to be multiplied by the appropriate value of time. For all schemes, we can see how introducing RT was beneficial thanks to reductions in AML overcompensating the addition of RT—even when premiums apply—, especially for service groups with high levels of AML (e.g. service 1).

Conclusions

Recovery time is the additional travel time that is built into a train service timetable over and above the minimum journey time necessary, often with the aim of reducing the probability of being late. Recovery time is widely used in railways in a number of countries but prior to this study there has been no investigation of the rail users' point of view. This paper summarises the findings of the first survey of rail users on the use of recovery time by train operators and their valuations of it. The paper also adds to the literature on late time valuations. The entire area, of late arrivals and means of reducing them, is significant since hundreds of millions of pounds are involved in financial compensations just in the railway industry in Great Britain and more generally there are appreciable impacts on the economic welfare of rail travellers.

The survey included a Stated Preference (SP) experiment aimed at exploring the trade-offs between travel time, recovery time and late time. While transport planning worldwide uses different valuations for all components of generalised journey time, there was no evidence prior to this study to indicate whether and to what extent recovery time should carry a valuation premium relative to in-vehicle time. The results of this work have been included in the Passenger Demand Forecasting Handbook (ATOC 2013) which represents the official railway industry guidelines in the UK.

The surveys reveal that most rail users are very supportive of the inclusion of recovery time in timetables. At the same time, a minority of users disapprove the use of recovery time as a tool to reduce lateness. Leisure travellers, followed by business travellers, are the most supportive. This can be explained by the infrequent nature of their journeys, and the associated importance of being on time for special occasions. Commuters are slightly less supportive presumably because they use the rail service often and, although they still care about being on time, adding extra time for every journey is a less appealing solution (this is especially the case for users with long commutes). These diverse preferences translated into a heterogeneous set of valuation estimates.

Our results show that the perceived recovery time carries a premium but only in some circumstances. Relative to in-vehicle time, perceived recovery time is valued between 1 and 3.5 times in-vehicle time, depending on journey purpose, distance, interchanges and awareness. However, if these results are to be applied for appraisal and forecasting purposes, the premium estimates should be evaluated relative to perception of late time. This is crucial as we have also found late time to carry a higher premium than the typical late time multipliers used in the UK, the context of the study. Our estimates show that 1 min of late time is valued at nearly 6 min of in-vehicle time (just above 5 for commuters), which is at the high end of the estimates found in the literature. Altogether, the recovery time multiplier is highly context-dependent and likely to be only slightly above 1 in many cases.

Controlling for the variability of late time improves the model fit and suggests the presence of risk-averse behaviour, but has a small influence of the estimates.

We provide an illustrative application of recovery time valuation and a guidance for how to use these in practice. The recommended appraisal application has been demonstrated using data from an actual scheme where recovery time was extended in the timetable of a series of rail services. In all cases, recovery time leads to a reduction in late time and an overall reduction in generalised journey time, concluding that schemes of this nature can be beneficial even when valuation premiums beyond the value of in-vehicle time are attached to recovery time.

Further research on valuation of recovery time would be highly desirable to build upon the findings of this first study. The outcomes of the study can help to improve the design of future stated preference experiments in this area, in particular the selection of trade-offs. For instance, widening the range of boundary levels and/or masking the focus of the survey (to avoid strong views leading to strategic bias) might help to reduce the non-trading behaviour observed in this context; new debrief questions could also help to further unpack the reasons for non-trading if it occurs Also, the evidence generated can now inform "priors" of future studies, contributing to building more efficient SP designs. However future SP experiments should also cover as an explicit variable different forms of delay-repay scheme given their increasing role in the industry. Future studies that limit the non-trading levels should also explore more advanced models to shed more light on the likely value distributions. Additionally, econometric analysis of ticket sales data, which is very common in rail industry, could be used to verify the findings by analysing routes where there had been changes in recovery time.

Authors' contributions MOC: Data Analysis and Modelling, Manuscript Writing and Editing. JS: Literature Search and Review, Manuscript Writing and Editing. MW: Preliminary work, SP design and Data collection, Manuscript Writing and Editing. FT: Preliminary work, Data collection and Editing. NH: Background information and data for case study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abrantes, P.A., Wardman, M.R.: Meta-analysis of UK values of travel time: an update. Transportation Research Part A: Policy and Practice **45**(1), 1–17 (2011)

ARUP, Institute for Transport Studies University of Leeds and Accent: Provision of market research for value of time savings and reliability. Phase 2 report to the Department for Transport. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470231/vtts-phase-2-report-issue-august-2015.pdf (2015)

Association of Train Operating Companies (ATOC): Passenger demand forecasting handbook version 5.1. Association of Train Operating Companies (ATOC), London (2013)

Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Ojeda-Cabral, M., Chintakayala, P., 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., Worsley, T.: New appraisal values of travel time savings and reliability in Great Britain. Transportation **46**, 1–39 (2017)

Bliemer, M.C.J., Rose, J.M.: Experimental design influences on stated choice outputs: an empirical study in air travel choice. Transportation Research A **45**(1), 63–79 (2011)

Fowkes, A.S.: Recent developments in Stated Preference techniques in transport research. In: PTRC conference, reprinted in Ortuzar J.D. (2000) Stated Preference modelling techniques, pp. 37–52. PTRC, London (1991)

Hess, S., Rose, J.M., Polak, J.: Non-trading, lexicographic and inconsistent behaviour in stated choice data. Transportation Research Part D: Transport and Environment **15**(7), 405–417 (2010)

Hess, S., Daly, A., Dekker, T., Ojeda-Cabral, M., Batley, R.: A framework for capturing heterogeneity, heteroskedasticity, nonlinearity, reference dependence and design artefacts in value of time research. Transportation Research Part B: Methodological **96**, 126–149 (2017)

Li, Z., Hensher, D.A., Rose, J.M.: Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence. Transportation research part E: logistics and transportation review **46**(3), 384–403 (2010)

Liu, X., Polak, J.W.: Nonlinearity and specification of attitudes toward risk in discrete choice models. Transp. Res. Rec. **2014**(1), 27–31 (2007)

Mackie, P., Wardman, M., Fowkes, A.S., Whelan, G., Nellthorp, J., Bates J.J.: Values of travel time savings in the UK. Report to Department for Transport (2003)

McFadden, D., Train, K.: Mixed MNL models for discrete response. Journal of applied Econometrics **15**(5), 447–470 (2000)

OECD/International Transport Forum: Valuing convenience in public transport: roundtable summary and conclusions. Discussion Paper 2014-02. OECD, Paris (2014)

Pachl, J.: Railway operation and control. VTD Rail Publishing, Mountlake Terrace (2002)

Palmqvist, C.W., Olsson, N.O.E., Hiselius, L.: Delays for passenger trains on a regional railway line in southern Sweden. International Journal of Transport Development and Integration **1**(3), 421–431 (2017)

Parbo, J., Nielsen, O.A., Prato, C.G.: Passenger perspectives in railway timetabling: a literature review. Transport Reviews **36**(4), 500–526 (2016)

Passenger Focus: Passenger focus response to the Office of Rail Regulation Consultation "Periodic review 2013: consultation on incentives" (2012)

Schittenhelm, B.: Planning with timetable supplements in railway timetables. In: Proceedings from the annual transport conference, Aalborg University (2011)

Senna, L.A.D.S.: The influence of travel time variability on the value of time. Transportation **21**, 203–228 (1994)

Significance, VU University Amsterdam and John Bates: Value of time and reliability in passenger and freight transport in the Netherlands. Report for the Ministry of Infrastructure and the Environment, The Hague (2012)

Teklu, F., Wardman, M., Segal, J., Dix, M., Condry, B.: Demand impacts of recovery time on railway timetables. In: Conference paper presented at the European Transport Conference (2012)

Train, K., Weeks, M.: Discrete choice models in preference space and willingness-to-pay space, Ch. 1. In: Alberini, A., Scarpa, R. (eds.) Applications of simulation methods in environmental resource economics, pp. 1–17. Springer, Dordrecht (2005)

UIC: Timetable recovery margins to guarantee time keeping—recovery margins. International Union of Railways, UIC CODE 451-1 OR (2000)

Wardman, M., Batley, R.: Travel time reliability: a review of late time valuations, elasticities and demand impacts in the passenger rail market in Great Britain. Transportation **41**(5), 1041–1069 (2014)

Wardman, M., Ibáñez, J.N.: The congestion multiplier: variations in motorists' valuations of travel time with traffic conditions. Transportation Research Part A: Policy and Practice **46**(1), 213–225 (2012)

Wardman, M., Chintakayala, V.P.K., de Jong, G.: Values of travel time in Europe: review and meta-analysis. Transp. Res. Part A Policy Pract. **94**, 93–111 (2016)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Manuel Ojeda-Cabral is a Senior Research Fellow at the Institute for Transport Studies, University of Leeds. His research is focused on the economic appraisal of public sector projects and policies and, in particular, cost-benefit analysis and the valuation of non-market goods such as time, accessibility, the environment and urban realm. Most of his work falls within the remit of the transport sector, where Manuel works closely with major public sector and industry organizations.

Jeremy Shires is a Senior Research Fellow and Market Research Manager at the Institute for Transport Studies (University of Leeds), with over 25 years of experience in transport research. His main research interests lie in transport policy, stated preference design, appraisal and forecasting within the public transport field. The majority of his work has been in the bus and rail sectors but he maintains strong interests in the fields of disabled transport provision and new technologies.

Mark Wardman is an independent researcher and Visiting Professor at the Institute for Transport Studies, University of Leeds. Mark has been involved in transport research for over 35 years, and a particular feature has been the analysis of the demand for rail travel. He has published extensively in this area and his research has impacted significantly on the demand forecasting methods and parameters that are widely used in the railway industry in Great Britain.

Dr Fitsum Teklu is a Market Director at SYSTRA who also works as project director and technical lead for transport model development, policy and revenue advice contracts. He has led the development of rail, multimodal, highway, and public transport models, supported by appropriate calibration techniques. Fitsum has used the models and analysis to support financial and economic appraisal and consequent decision making for both public and private sector clients.

Nigel Harris has been the Managing Director of the Railway Consultancy since 1995, after 8 years in strategic and network planning at London Underground. Whilst retaining academic interests at several universities, Nigel concentrates on the practical applications of demand forecasting and service planning, and he has worked for every British train operator except one, as well as in 40 other countries around the world.