Article:

Wardman, M and Batley, R (2014) Travel time reliability: a review of late time valuations, elasticities and demand impacts in the passenger rail market in Great Britain. Transportation, 41 (5). pp. 1041-1069. ISSN 0049-4488

https://doi.org/10.1007/s11116-014-9526-5

TRAVEL TIME RELIABILITY: A REVIEW OF LATE TIME VALUATIONS, ELASTICITIES AND DEMAND IMPACTS IN THE PASSENGER RAIL MARKET IN GREAT BRITAIN

Mark Wardman and Richard Batley

Institute for Transport Studies, University of Leeds

CORRESPONDING AUTHOR

Mark Wardman
Institute for Transport Studies, University of Leeds
Leeds, LS2 9JT
United Kingdom
Telephone: +44 (0) 113 343 5349
Fax: +44 (0) 113 343 5334
E-mail: M.R.Wardman@its.leeds.ac.uk

ABSTRACT

This paper provides an extensive review and reconciliation of British and European evidence relating to the value of, and demand responses to, rail reliability. In particular, we compare the elasticities implied by Stated Preference (SP) valuations of late time with directly estimated lateness elasticities. We find that the implied lateness elasticities are substantially greater than those directly estimated. A possible explanation for this is that lateness has been over-valued, but more sobering explanations would be to suggest that, whilst rail travellers dislike unreliability, they may be unwilling or unable to reduce their rail travel in response to experiences of poor performance, or else conventional economic approaches to deducing elasticities are not appropriate. The findings have been used to update the recommendations of the UK rail industry's Passenger Demand Forecasting Handbook.

KEYWORDS

Rail reliability, valuations of late time, lateness elasticities.

ACKNOWLEDGEMENT

The authors are grateful for the funding received from the Association of Train Operating Companies and the Office of the Rail Regulation to support this research and would like to express their thanks to John Segal and Fitsum Teklu of the MVA Consultancy for their support and advice. The views expressed here are those of the authors and do not necessarily reflect those of the sponsors. Thanks are due to the comments of four anonymous referees.

1. INTRODUCTION

Time savings have long been accepted as having important benefits to travellers and also measurable impacts on demand. Whilst *variability*[1] in travel time (which has attracted the term *reliability* in the literature) has also for some time been recognised as the source of considerable inconvenience to travellers and a deterrent to travel, its treatment in real-world forecasting and appraisal has lagged far behind that for travel time savings. Historically this has been due to difficulties in both valuing reliability and forecasting the benefits of its improvement. We are now at a point, however, where a considerable amount of attention has been paid to the development of the theoretical representation, methodological treatment and empirical estimation of the benefits of improved reliability. Despite this accumulation of the research base, and recognition of the importance of reliability to travellers, the inclusion of reliability in the appraisal of transport schemes, policies and investments remains at best patchy but generally negligible in most countries.

A notable exception to this is the railway industry in Great Britain, where a procedure has been in place since 1986 to forecast the impacts of changes in reliability on passenger rail demand. This procedure is based around the concept of mean lateness on arrival (at the destination station), the valuation of mean lateness in units of scheduled travel time[2], the forecasting of any demand response to a change in mean lateness (via the 'lateness multiplier' and travel time elasticity), and practical operationalisation through a wealth of information on actual train running times. This procedure encompasses not only late running of services relative to the timetable, but also cancellations through a formula which converts service cancellation into 'deemed' minutes of lateness.

This forecasting procedure, set out in the railway industry's Passenger Demand Forecasting Handbook (ATOC, 2013), yields *implied* elasticities of demand with respect to late time. An alternative procedure,

[1] In this context, we are primarily concerned with random (e.g. as might be caused by an unforeseen incident) as opposed to systematic (e.g. as might be reflected by a longer travel time during peak hours relative to off-peak) variability.

[2] The so-called lateness multiplier expresses late arrival time in equivalent units of scheduled travel time.

preferable in principle, would be to estimate the late time elasticities *directly*, as a result of quantitative analysis of the relationship between demand and, amongst a range of other things, mean lateness or some other suitable representation of reliability. There is now an emerging body of evidence in this area.

Reviews of reliability valuations are present in the literature (Tseng, 2008; Bates, 2010; Carrion and Levinson, 2012), but the present paper adds to these with a significant amount of British evidence not otherwise in the public domain. What is a more original contribution is the review of evidence on directly estimated reliability elasticities, again much of which is not in the published literature. This places us in the unique position of being able to compare *directly* estimated reliability elasticities with those *implied* using late time multiplier evidence. In summary, therefore, the objectives of this paper are three-fold. First, we review a large amount of valuation evidence. Second, we review emerging elasticity evidence. Third, we endeavour to reconcile these two distinct sources of evidence.

Concerns over the quality of unpublished evidence, as used in this review, are sometimes raised. It is not practical here to discuss the quality of each and every study, and such a discussion would in any event be subjective. However, we would make the following points which justify our approach on this matter. First, of the 21 studies covered here, 5 are journal papers, 3 are conference papers, 3 are published on the web and 10 are, as far as we are aware, unpublished; thus the majority of our evidence base has been published in some shape or form. Second, we have previously conducted the most extensive meta-analyses ever of values of travel time (Abrantes and Wardman, 2011), time-based elasticities (Wardman, 2012) and price elasticities (Wardman, 2014), and only in the latter was unpublished evidence found to provide different results (and even then the discrepancy was minor). Third, reliance on published evidence would severely restrict the extent of the evidence base. Fourth, the unpublished evidence emanates largely from industry research undertaken by respected research consultants and using best practice methods.

The structure of the paper is as follows. Section 2 discusses key background issues, whilst section 3 provides a review of reliability valuations, and section 4 covers directly estimated elasticity evidence. A synthesis of the two sets of evidence is provided in section 5, with concluding remarks in section 6.

2. BACKGROUND

There are a number of important background issues we here discuss. These include the treatment of reliability within models estimating its effect on utility, choice and demand, the importance of reliability to train travellers and its treatment in forecasting, and the reasons why directly estimated and implied elasticities might differ.

2.1 Theoretical Issues

Reliability influences rail travellers' wellbeing and their choices. In modelling these choices and inferring the value of reliability, the transportation literature has drawn analogy to the extant literature on the economics of money risk. According to this analogy, the individual traveller is, in the face of travel time variability, assumed to choose the travel option that maximises expected utility[3] (von Neumann and Morgenstern, 1947). In this context, the mathematical representation of expected utility has given rise to three dominant approaches, as follows.

The first approach, referred to as the mean-variance approach, emanates from portfolio analysis (Tobin, 1958, 1965; Markowitz, 1959), a branch of the economics literature. According to this approach, which was first applied in the reliability context by Jackson and Jucker (1981), expected utility is approximated by the first and second moments of the utility distribution over the variation in travel times.

$$U_i = \gamma \overline{T}_i + \delta \sigma_i + \ldots\ldots \qquad \text{for all } i \in M \qquad (1)$$

[3] Transport researchers have also shown interest in non-expected utility paradigms such as Prospect Theory (Kahnemann & Tversky, 1979), but expected utility (it would seem) continues to prevail in both the economics and transportation literatures.

where M is a set of discrete travel choices (such as mode or destination), $\overline{T}_i$ is the mean travel time for alternative i and σ_i is the standard deviation of travel time. A frequently cited metric emanating from (1) is the so-called 'reliability ratio' (RR), defined as the ratio δ/γ. In practical transport modelling, the mean-variance approach has found particular favour in the representation of reliability on road. However, a variant on this approach, particularly relevant to rail travel, involves the specification of mean delay and standard deviation of delay in addition to scheduled travel time.

Whereas the mean-variance approach is applicable to the contexts of both money or time risk, the second approach, referred to as the scheduling approach (Vickrey, 1969; Noland and Small, 1995), appeals specifically to the context of time risk. The approach is framed around an interest in how travellers choose their departure time when seeking to arrive at their destination by a 'Preferred Arrival Time' (PAT), and gives rise to the following expected utility function:

$$U_i = \varphi E(T_i) + \kappa E\left(\max\left(SDE_i, 0\right)\right) + \mu E\left(\max\left(SDL_i, 0\right)\right) + \tau E(D_i) + \ldots\ldots \quad \text{for all } i \in M \qquad (2)$$

where T_i is the travel time for alternative i, and $E\left(T_i\right)$ is its expectation across the travel time distribution; SDE_i is Schedule Delay Early, the amount of time the actual arrival is before the PAT, and $E\left(SDE_i\right)$ is again its expectation; SDL_i is Schedule Delay Late, the amount of time the actual arrival is after the PAT, and $E\left(SDL_i\right)$ is its expectation; D_i denotes a lateness dummy variable (=1 if $SDL_i > 0$, and =0 otherwise), and the expectation $E(D_i)$ can be interpreted as the probability of being late. Like the mean-variance approach, the scheduling approach has been applied particularly to reliability on road, although applications to public transport are also evident in the literature; we will comment further on this in due course.

The third approach, referred to as the mean lateness approach, entails the expected utility function:

$$U_i = \alpha AML_i + \beta ST_i + \ldots\ldots \quad \text{for all } i \in M \qquad (3)$$

where, amongst other things, AML_i is average minutes of lateness relative to the *public transport* schedule at the destination station for alternative i, and ST_i is the scheduled travel time. The ratio $w_{AML} = \alpha/\beta$ is the value of mean lateness in units of scheduled travel time, that is, the 'lateness multiplier' referred to in section 1. The theoretical credentials of this approach are weaker than the other two[4], and its origins lie in the need for a practical approach of valuing reliability that, whilst emulating the likes of the mean-variance approach, is more applicable to the context of scheduled public transport services.

In empirical studies, combinations of approaches are possible, as investigators search for the best explanation of the data. There are a number of significant relationships that fall out of this discussion:

- Given that there will be some inconvenience or penalty involved in arriving late, the 'values' or 'multipliers' of SDL (i.e. μ/φ from (2), in units of actual travel time) and mean lateness (i.e. $w_{AML} = \alpha/\beta$ from (3), in units of scheduled travel time) can be expected to exceed one.
- If the PAT is interpreted as the scheduled arrival time of a train service, then SDL (from (2)) will be equivalent to late time (from (3)). On the other hand, if the train is late but arrives before a traveller's PAT, the amount of lateness will be less than the amount of SDL. Indeed, for some travellers, lateness might move them closer to their PAT. We would therefore expect the multiplier for mean lateness to be less than that for SDL, i.e. $w_{AML} < \mu/\varphi$.
- The value of SDE is not only expected to be somewhat less than for SDL, but is expected to be less than one, on the grounds that arriving early is less onerous than travel time; this is to say, it would generally be possible to pursue the same activities in the early time as when travelling but for some travellers there will be opportunities to spend the time more usefully.

[4] See Batley and Ibáñez (2013) for a more detailed comparison of the theoretical properties of the three approaches.

- Under the specific circumstances of continuously variable departure times, which admittedly do not generally hold for rail travel and are only approximated in a few circumstances, $\kappa E\left(\max\left(SDE_i,0\right)\right)+\mu E\left(\max\left(SDL_i,0\right)\right)$ is closely approximated by $H\left(\kappa,\mu\right)\cdot\sigma$ for a wide range of travel time distributions (Bates et al., 2001; Fosgerau and Karlström, 2009). Indeed, the reliability ratio can be expressed as a function of the SDE multiplier and the ratio between the SDE and SDL multipliers; we will discuss this further in due course.

The discussion above has related to the impact of reliability on individuals, as in the world of choice modelling. However, there is another world that economists inhabit which involves the analysis of the outcomes of many individuals' choices. This is the world of demand analysis and elasticities[5]. Conveniently, the concept of mean lateness introduced in (3) readily translates to the context of station-to-station flows, which is the desirable level at which rail demand analysis should operate and is the convention in Great Britain. Instead of a late time valuation, such analysis elicits a late time elasticity indicating the proportionate change in demand after a proportionate change in mean lateness. Variants might involve the specification of the standard deviation of lateness or instead the mean and standard deviation of travel times. What is not possible though is the estimation of elasticities to SDL, since this is an individual-specific term that cannot be replicated at the station-to-station flow level.

2.2 The Importance of Train Reliability

From a number of perspectives, train reliability can be seen to be important for all the key players in the rail market. Within the British railway network, reliability represents a formal element of the so-called 'Schedule 8 performance regime', which is a central tenet of franchised train operating companies' (TOCs) track access contracts. Schedule 8 seeks to incentivise both the TOCs and the infrastructure provider (Network Rail) to minimise lateness and the cancellation of train services. This is operationalised through a system of compensation payments which results in significant financial flows

[5] Of course, discrete choice models based on appropriate choice contexts yield demand elasticities as well as valuations for reliability. Nonetheless, in this context they have been almost exclusively used for valuation.

and a key driver in these payments is the lateness multiplier w_{AML}. In addition to these regulated payments, a further financial flow is compensation to passengers through the provisions of the Passengers' Charter and Delay-Repay schemes.

The importance of on-time arrivals is clear in survey evidence. Passenger Focus (2012) reported it to be the most important driver of passenger satisfaction "by a considerable margin" and the second priority for improvement after Value for Money (VfM). The 'McNulty Report' into VfM in the UK rail industry (Department for Transport and Office of Rail Regulation, 2012) stated: *"There is strong evidence to suggest that the historic high levels of customer satisfaction…achieved throughout 2010 have been largely driven by the equally historic high level of punctuality and system performance attained concurrently. Indeed, it seems clear that train performance is the largest single driver of customer satisfaction"* (p265). This qualitative evidence is confirmed in the findings of quantitative research. In economic parlance, an attribute's importance is defined as how much travellers are willing to pay to improve it. Abrantes and Wardman (2011) reported the w_{AML} implied by their meta-model of a large amount of UK evidence to vary from 3.7 for urban journeys to around 1.4 at 100km. These valuations contrast with implied multipliers for walk and wait time of around 1.7 for urban journeys and around 1.2 for 100km journeys. Given that walk and wait time are traditionally regarded as major impediments to public transport use, these valuations of late time can be regarded as particularly significant.

2.3 The Treatment of Reliability in the UK Rail Industry

The Passenger Demand Forecasting Handbook (PDFH) was not only a pioneer in the field but remains unique in railway administrations across the world in setting out a demand forecasting framework, largely elasticity driven, and more importantly recommending a set of regularly updated parameters based on a distillation of the best empirical evidence to populate that framework.

The first edition appeared in 1986 and would seem to be one of the earliest, if not the earliest, treatments of reliability in a demand forecasting context. This treatment of reliability entails an extension to its forecasting of timetable-related service quality changes under certainty, which is based on a composite term referred to as Generalised Journey Time (GJT). This takes the form:

$$GJT = ST + \lambda H + \varpi I \tag{4}$$

where ST is the scheduled station-to-station travel time, including any time spent interchanging, H is service headway and I is the number of interchanges. The weights attached to headway (λ) and interchange (ϖ) convert these variables into equivalent amounts of scheduled travel time[8]. The proportionate change in the volume of demand (V) between the new and base situations is forecast as:

$$\frac{V_{new}}{V_{base}} = \left(\frac{GJT_{new}}{GJT_{base}} \right)^{\eta_{GJT}} \tag{5}$$

where η_{GJT} is the elasticity of demand with respect to GJT. Much evidence has long existed for η_{GJT} obtained from econometric analysis of station-to-station ticket sales data (Wardman, 2012).

In principle, GJT in equation (4) could be enhanced with the addition of mean lateness (L), and the GJT elasticity then re-calibrated. Unfortunately, there is very limited evidence on this re-calibrated elasticity (Batley et al. (2008) is an isolated example), and the approach adopted in practice has been to forecast changes in reliability, represented by mean lateness, as if it were an equivalent change in GJT:

$$\frac{V_{new}}{V_{base}} = \left[1 + \frac{w_{AML}(L_{new} - L_{base})}{GJT_{base}} \right]^{\eta_{GJT}} \tag{6}$$

This is approximately true for late time changes that are a small proportion of GJT. As is apparent from equation (6), the valuation of late time (w_{AML}) is a critical parameter. Whilst the framework has not changed over time, the recommended w_{AML} multipliers have, as have the recommended η_{GJT}.

[8] Given that timetable patterns often vary across the day, GJT is calculated in 15 minute time intervals and is weighted according to a profile of desire departure times. Note that in practice the λ and ϖ weights are not (as represented here for simplicity) constant.

The first edition of PDFH in 1986 recommended a w_{AML} of 2.5 based on the work of Benwell and Black (1984). The second edition, issued in 1989, retained the w_{AML} of 2.5 for business and leisure, but now supported by the findings of MVA (1989), recommended revised multipliers for commuting, specifically a w_{AML} of 1.25 as well as an additional multiplier of 3.2 for the standard deviation of lateness. This was, as far as we are aware, the first time rail appraisal practice covered the variability of late time. PDFH3 (1997) reverted to a w_{AML} of 2.5 for all passengers due to concerns raised by Jones et al. (1995) about the high weighting attached to the standard deviation of lateness. PDFH4 (2002) took on board the pioneering work by Bates et al. (2001) and increased w_{AML} to 3 for all passengers but also suggested that sensitivity analysis might also use a w_{AML} of 6.5 for airports, 6.1 and 4.2 for long distance 'high speed' full fare and reduced fare paying passengers, and 2.5 for all others. Since the 2002 update, a significant amount of fresh evidence on w_{AML} has emerged. This is the subject of the review reported here, and the basis of the most recent PDFH update in 2013 (PDFH5.1) as set out in ATOC (2013). Alongside this is new evidence indicating the direct influence of reliability changes on rail demand.

Turning to η_{GJT}, the other critical parameter in the forecasting equation (6), PDFH recommended a figure of around -0.9 until 1997, when PDFH3 allowed it to vary (between around -0.5 and -1.1) according to the level of competition from other modes and the quality of the rail service. After some simplifications introduced in 2005, the most recent PDFH update of 2013, which for GJT was largely driven by the meta-analysis of Wardman (2012), clarified that the recommended η_{GJT} were explicitly long run effects.

2.4 Implied Elasticities

Given that reliability is regarded as a key aspect of train travel by passengers, and the significant premium attached to it relative to in-vehicle time, it is very easy to be drawn into the view that the elasticity to travel time reliability will be large. Indeed, since the elasticities inherent in equation (6) are rarely considered, this view has not been exposed to challenge. By way of illustration, and to facilitate comparison with *directly* estimated elasticities in section 5 to follow, we can deduce[6] the *implied* late time elasticity (η_{AML}) from PDFH's recommended η_{GJT} as:

$$\eta_{AML} = \eta_{GJT} \frac{w_{AML} AML}{GJT} \tag{7}$$

As illustrations of the likely range, we can use two contrasting sets of values. The highest PDFH5.1 (ATOC, 2013) recommended η_{GJT} is -1.35 for long distance London flows which also have the highest mean lateness of 5.2 minutes. If we take the traditional w_{AML} of 3 which prevailed from 2002, then for a GJT of 120 minutes, the implied η_{AML} is -0.18. This falls to -0.11 for non-commuters within Greater London, who have the lowest η_{GJT} of -0.9, a mean lateness of 1.2 and a typical GJT of 30 minutes. Very short (long) and unreliable (reliable) flows will tend to have largest (smallest) implied η_{AML} , with the amount of variation moderated to the extent that longer distance flows tend to be less reliable.

2.5 Divergence between Directly Estimated and Implied Elasticities

---

[6] Suppose late time were entered into GJT to create an enhanced GJT (EGJT). We would then have:

$$EGJT = GJT + w_{AML} AML$$

and for forecasting purpose we would have:

$$V = EGJT^{e}$$

The implied elasticities to GJT and lateness would be:

$$\eta_{\ GJT} = e \frac{GJT}{EGJT}; \quad \eta_{AML} = e \frac{w_{AML} AML}{EGJT}$$

Rearranging terms:

$$e = \frac{EGJT}{GJT} \eta_{GJT} = \frac{EGJT}{w_{AML} AML} \eta_{AML}$$

It follows that:

$$\eta_{AML} = \eta_{GJT} \frac{w_{AML} AML}{GJT}$$

There are several reasons why any direct evidence on the demand response to reliability changes might differ from the implied demand effects using equation (6).

The w_{AML} term in equation (6) comes from Stated Preference (SP) evidence. Arriving late is a contentious issue and one that might attract strategic bias in SP responses, particularly given that it will often be apparent to respondents that the questionnaire is centred around the importance of late time. We might therefore expect w_{AML} values to be too high, whereupon the implied elasticities will be too large; indeed, the literature contains some estimates of w_{AML} which might be regarded as extreme.

Directly estimated elasticities will, to a large extent, incorporate 'market imperfections' such as individuals behaving in a habitual manner and, more importantly, lack of awareness of real-world changes in reliability. These will tend to dampen the elasticities compared to the 'perfect information' scenarios of SP. The same is true if respondents do not fully account for all of the real world constraints on behavioural change when undertaking SP exercises.

Whilst equation (6) is an approximation, this need not concern us unduly since changes in reliability will tend to be small relative to GJT. What is potentially a greater concern is that the forecasting method used of converting late time changes into equivalent units of GJT might not be empirically justified, not least because the average w_{AML} estimated in SP studies (of rail users) is not necessarily the same as the value at the margin that drives behavioural change. That is to say, once a traveller is a rail user then arriving on time might become a very important issue, but it might not be pivotal to inducing mode switching in non-rail users. On the other hand, directly estimated demand elasticities will tend to reflect the marginal effects and hence be lower than those implied by equation (6).

Finally, equation (6) essentially forces the individual to accept any changes in reliability and forecasts on the basis of this. However, travellers might mitigate the consequences of poorer performance, say by travelling on an earlier departure to get to work on time or by using another route or operator. Given that such mitigations must, by definition, be preferable to 'being forced' to face the worsened reliability, the demand effect can be expected to be less than equation (6) would imply.

In summary, these reasons lead us to expect, and more than just on balance, that directly estimated lateness elasticities would be lower than implied elasticities.

2.6 Experience Elsewhere

Our primary focus here is on the treatment of reliability in forecasting and appraising rail demand in the rail industry in Great Britain, which is unique in having had procedures to do this for many years. Nonetheless, other relevant evidence, experience and practice should be considered.

We note that the UK government's appraisal guidance, quite separate from PDFH's demand forecasting recommendations, specifies a reliability ratio of 0.8 for car and 1.4 for all public transport modes, defined in the latter case as the ratio of the marginal utilities of standard deviation of late time and mean late time. Nonetheless, benefits from reliability improvements are not included within core scheme benefits. In Sweden, public transport late time has a multiplier of 3.5, whilst a reliability ratio of 0.9 applies to car. New South Wales in Australia uses a late time multiplier of 3.7 for public transport, whilst in Denmark it is 2. In the USA, the official reliability ratio ranges from 0.8 to 1.1.

Whilst these valuations indicate how much travellers benefit from improvements in performance, they are not behavioural indicators. Dutch railways use passengers' judgements, on a 1-10 scale, of punctuality and reliability of trains that feed into a general rating of how good the train service is. In this case, forecasting is based on changes in the general customer judgement, to which an elasticity parameter is applied. However, we are not aware of other railway administrations that account for the impact of reliability on demand in anything like the formal manner that has long been conducted in Great Britain.

3. VALUATION EVIDENCE

The valuation of travel time variability has relied upon SP evidence, on the grounds that it is difficult to identify real world situations involving trade-offs between reliability and appropriate numeraire such as time or cost in which units valuations are invariably expressed. Even then, travellers might not always be aware of these trade-offs, and measuring reliability in a manner that reflects travellers' perceptions has therefore proved challenging.

The stochastic nature of travel time variability means that its presentation to respondents in SP exercises requires careful thought, ideally with suitable exploratory research and certainly thorough piloting. Some studies have presented a number of trips, usually five or ten, with different amounts of travel time or late time, each with the same likelihood of occurrence; a variant of this is the 'clock face' format (Bates et al., 2001). An alternative approach is to say that trains will be X minutes late one in Y times, although a concern here is that, unless explicitly stated, respondents might not interpret the remaining trains as on time. There has even been experimentation with presenting a frequency distribution of travel times. For a detailed discussion of such issues, see Hollander (2009).

3.1 Review of UK Evidence

We distinguish here between various types of valuations that have been estimated, although some studies provide more than one type of valuation, either from separate models or from a single model specification. Whilst some studies recover reliability values in monetary units, our primary concern has been with time multipliers, which account for the majority of the evidence. Indeed, it seems more natural to value and forecast reliability in time units, and indeed this is how it is used in the railway industry in Great Britain and elsewhere.

We have not reported studies that yield values that are not readily interpreted or cannot contribute to the PDFH forecasting approach (6), such as valuations of improving from 'quite reliable' to 'very reliable' services, or from 'X% on time arrivals to Y%', or the value of maximum delay (London Transport, 1983; MVA, 1985; MVA et al., 1987; Steer Davies Gleave, 1994; Burge et al., 2010).

We here cover 15 UK studies, although we have taken the opportunity to include three relating to bus and one to car given that bus is in many respects similar to train and the car study has novel aspects. We do not provide a detailed account of each study but summarise the key features in Table 1 which indicates the measure of reliability employed (Mean Lateness, Standard Deviation of Lateness, Schedule Delay Late or the Reliability Ratio), the means of presentation, the mode, and the journey purpose. Whether the study was a journal article, published on the web, a conference paper or unpublished is also given along with details of the number of respondents and the form of SP exercise used.

The most common means of presenting reliability in the earlier studies was to state that there would be a late arrival of some amount say one in five journeys. This approach is still apparent in recent studies but the presentation of, say, five trips with different travel times has become more common. The most commonly estimated valuation is mean late time at destination, and no doubt a contributory factor here is its use in forecasting by PDFH.

There are a number of notable features of individual studies in Table 1. Firstly, Lu et al. (2008) experimented with the use of 'cheap talk' wording to reduce the incentive to strategic bias that might lead to exaggerated willingness-to-pay values, by reminding respondents of the possible financial consequences of improved train services. The late time values fell by, on average, around a quarter in the presence of cheap talk[7]. Secondly, the literature tends to focus on the reliability of arrivals at the destination but WS Atkins (1997) and Batley et al. (2006) were innovative in exploring reliability in

[7] Usually, cheap talk works through in terms of cost sensitivity, which would leave time based multipliers unaffected. In this study, the cheap talk effect was interacted with the late time (and rolling stock) coefficient.

waiting time for bus and rail respectively. These studies demonstrate that, as expected, it is not just arrivals at the destination that are important but that variability at the origin is an issue. Thirdly, a considerable amount of attention has focussed on how best to present to respondents the inherently complicated issue of travel time variability. After extensive exploratory research, the Faber Maunsell (2003) study arrived at the challenging conclusion that respondents related best to a graphically presented probability distribution of travel times rather than the more conventional means of presentation such as five journeys. Fourthly, it is noticeable that the one study based on mode choice (TPA, 1992), where reliability was not the main focus of the SP exercise and hence less likely to attract strategic responses, recovers a late time multiplier very much at the low end of the range reported. Finally, the pioneering Bates et al. (2001) study, upon which so much emphasis has been placed by the rail industry in Britain, did not actually estimate time multipliers but instead imported values of time from review evidence (Wardman, 1998) to convert valuations estimated in monetary units into equivalent time terms. A similar approach was adopted in Benwell and Black (1984) and MVA (2000).

Taking the valuations of mean late time from Table 1 where it is the only term estimated, the average multiplier for commuting is 3.9 across three observations, and 6.3 and 4.3 respectively for leisure and business each for two observations. For trips defined separately as non-commuting, the average multiplier is 2.4 across three observations. Too much should not be read into the variations by purpose due to the small sample size, but taken together the average is 3.8, which is somewhat larger than the central figure of 3 used in PDFH.

As for the reliability ratio, it averages 0.91 for car and 1.07 for public transport, broadly in line with the Department for Transport's recommendations. With regard to the reliability ratio, Bates et al. (2001) concluded that: *"Values around 1.3 appear plausible for car travel; somewhat higher values may be appropriate for scheduled public transport, but values above 2 are unlikely"* (p228). The results here are lower than their recommendations.

Table 1: Summary Valuation Table

Table 1: Summary Valuation Table

Study	Source	Mean Late (α/β)	SD Late	SDL (μ/φ)	RR (δ/γ)	Sample and Method	Presentation	Mode	Purpose
Benwell and Black (1984)	Conference	2.5				384 individuals ranked 7 alternatives	10 trips	Rail	Non Commute
MVA (1989)	Unpublished	1.2 2.0	1.6 1.3			406 individuals ranked 10 alternatives	1 in X	Rail	Commuter Leisure
MVA (1991)	Unpublished		4.8 2.7 6.4 4.3 4.5 5.9 2.7			241 individuals ranked four sets of four alternatives	1 in 5	Rail	Business Commute Leisure <50 miles 50-140 miles (Full) 50-140 miles (Red) Over 140 miles
TPA (1992)	Unpublished	1.7				328 individuals choose between car and rail 8 times	1 in 5	Rail & Car	Non Commute
Black and Towriss (1993)	Unpublished				0.63 0.51 1.02 0.55 0.79	606 individuals making 12 within mode pairwise choices.	1 in 5	Rail Bus Car Car Car	All All EB Commute VFR
WS Atkins (1997)	Unpublished				1.4 (0.9)[1] 1.1 (0.8) 1.7 (0.6)	187 individuals making 9 within mode pairwise choices	1 in 5 and Range	Bus	Commute Shopping Other
Bates et al. (2001)	Journal	5.5 3.9 1.7 1.9 1.2 1.3 4.9 0.6 1.0 4.2 1.0		4.8 2.4 1.9 1.2 1.8 1.3 8.0 0.6 2.4 2.3 0.8		451 individuals making 6 within mode pairwise choices	'Clock' of 10 arrival times	Rail	Great Western EB Great Western Non-EB Connex Commute Connex Leisure Northern Spirit EB & Commute Northern Spirit Leisure Northern Spirit Airport Central Trains EB & Commute Central Trains Leisure Virgin Trains EB & Commute Virgin Trains Leisure

Study	Source	Mean Late (α/β)	SD Late	SDL (μ/φ)	RR (δ/γ)	Sample and Method	Presentation	Mode	Purpose
MVA (2000)	Unpublished	5.7 3.8 4.0	1.3 0.6 2.4			1073 individuals making 10 within mode pairwise choices	'Clock' of 8 arrival times	Rail	Inner Suburban Outer Suburban Inter City Leisure
Faber Maunsell (2003)	Unpublished				1.30	188 individuals making 6 within mode pairwise choices	Prob Dist[n]	Car	All
Batley et al. (2006)	Journal	2.70 (2.26)[2] 3.22 (2.72) 5.30 (3.53) 1.80 (1.96) 2.10 (2.11) 1.88 (2.23)			1.30 1.39 2.80 1.54 2.39 2.36	2395 individuals making 5 within mode pairwise choices	5 trips	Rail	Short EB Short Commute Short Leisure Long Business Long Commute Long Leisure
Hollander (2006)	Journal			2.75[3]		244 individuals making 9 within mode pairwise choices	5 trips	Bus	Non EB
Faber Maunsell and Mott MacDonald (2007)	Web	5.30 2.60 2.72 2.31				779 individuals making 8 within mode pairwise choices	1 in 5	Rail	All All Commute Non Commute
Lu et al. (2008)	Journal	3.27[4] 5.24 5.19				1222 individuals making 9 within mode pairwise choices	1 in 5	Rail	EB Commute Leisure
Steer Davies Gleave (2008)	Unpublished			2.59 4.81 3.66		1378 individuals making 9 within mode pairwise choices	5 trips	Bus	EB Commute Leisure
MVA and ITS Leeds (2012)	Conference	5.44 3.67 7.30				1013 individuals making 9 within mode pairwise choices	5 trips	Rail	EB Commute Leisure

Notes: Where more than one value is reported, they were obtained from the same model. [1] Reliability ratio for wait time in brackets. [2] Figures in brackets are mean lateness at boarding. [3] Ratio of SDL to combination of mean time and mean earliness. [4] Cheap talk values. Connex, Great Western, Northern Spirit, Central Trains and Virgin Trains are train companies.

3.2 Explaining the Valuation Evidence

The valuations vary quite a lot across studies, which is not surprising due to limited comparability. To try and make sense of what is a large and diverse evidence base, we chose to develop regression models. The dataset used, which is the mean lateness of valuations or values that we feel can approximate it, contains 41 observations. We have not included the Black and Towriss (1993), WS Atkins (1997) and Faber Maunsell (2003) studies since they relate to the reliability ratio, although in any event this is largely non-rail evidence.

We have taken the view that SDL provides an approximate estimate of mean late time, but in any event have included a variable in the modelling to denote where we use SDL as a proxy for lateness. Similarly, our view is that the standard deviation of lateness will be highly correlated with mean lateness. Hence in the MVA (1991) study we took the measure of mean lateness to be the standard deviation, but also use a variable to denote we have done this! For Bates et al. (2001), we have used the mean lateness values as reported, but have taken account of the fact that these may be correlated with other valuations. In an attempt to explain variation in mean lateness valuations, we have modelled the following features of the source data:

- Is the standard deviation of lateness used as a proxy for the mean lateness values?
- Is SDL used as the measure of mean lateness?
- Is there an additional delay type variable included in the reported model alongside the measure of mean lateness that we are using?
- Journey purpose
- How variability was presented
- Whether the journey was suburban, inter-urban or a mix
- Age of study (years from 2012 as well as ‘old’ or ‘new’)
- Whether the time multiplier was deduced from external value of time evidence
- Whether the value came from a mode other than rail.

Given there is only one airport observation, we did not include it in the dataset. The estimated models are reported in Table 2 below. These are linear-additive in form. The goodness fit measures are very modest, reflecting the disparate nature of the evidence, and is not surprising when dealing with a challenging valuation exercise, as is the case here. We have to assume that the relatively large amount of error is essentially random in nature; this assumption might seem strong, but at the same time we must accept that the dataset covers the extent of what evidence exists (in Great Britain, at least).

Table 2: Regression Analysis of Lateness Multipliers (W_{AML})

Model	**I**	**II**	**III**
Constant	3.92 (9.0)	3.84 (9.4)	3.97 (8.5)
Other Term	-2.03 (3.6)	-1.97 (3.5)	-1.75 (3.0)
Inter-Urban	1.16 (1.7)	1.17 (1.7)	0.83 (1.2)
Mix	1.08 (1.8)	1.13 (1.9)	0.92 (1.4)
Non Commute	-1.66 (1.6)	-1.82 (1.7)	-1.39 (1.2)
Adj R^2	0.27	0.25	0.18
Observations	40	40	41

Note: t-ratios in brackets.

Model I weights each study by one if it is rail and 0.5 otherwise. It has excluded one outlier which has a large standardised residual (the inner suburban multiplier of 5.7 in MVA (2000)) and retains those coefficient estimates with t-ratios exceeding one. The coefficients relating to the use of the standard deviation of lateness and SDL as proxy measures were far from significant. However, we observe that where there is an additional term to the mean lateness (Other Term), the mean lateness value is somewhat lower, as would be expected, and it is the most significant effect recovered.

Encouragingly, whether the time multiplier was deduced had no significant effect, and nor did the age of the study nor, surprisingly, the representations of reliability in the SP exercise. The only significant (at the 10% level) journey purpose effect was whether it was non-commuting (Non Commute). It seems reasonable that commuters have larger valuations of lateness; indeed it should be remembered that all the early studies of the value of travel time variability focussed on commuting trips with little recognition that it might be an issue for other journey purposes.

There are also larger valuations for longer distance trips (Inter-Urban). This is perhaps surprising given that lateness is more expected and hence 'excusable' on longer distance trips, but it might be a consequence of the proportion of business trips increasing with distance and such travellers being more sensitive to delay. Indeed, a larger multiplier for longer distances would offset the strong reduction in the implied late time elasticity for longer distance flows within the current PDFH approach. Values relating to both urban and inter-urban journeys (Mix) tend to be nearer the inter-urban values, although as with most coefficients the relatively low precision of estimation needs to be borne in mind.

Model II removes the weight on non-rail values with little difference in the results. Model III makes use of this weighting but does not remove the single outlier observation. Whilst the coefficient estimates do vary, there is a noticeable reduction in the goodness of fit and the t-ratios of coefficient estimates, thereby justifying removal of the outlier. Taking Model I as the preferred model, it implies a w_{AML} of:

- 5.08 for Inter-urban commuting
- 3.42 for Inter-urban non-commuting
- 3.92 for Suburban commuting
- 2.26 for Suburban non-commuting

We note, however, that there were no actual observations for inter-urban commuting and hence this effect should be treated with caution. Nonetheless, it should be noted that the implied w_{AML} generally exceed the recommended value of 3 in the 2009 edition of PDFH.

In calculating these multipliers, we have taken the definitions of inter-urban and suburban as used in the original studies. This means that there is no perfect definition of the segments, but non-London inter-urban flows typically involves distances greater than 20 miles, whereas London and South East flows may involve distances of up to 50 miles (extending to the old Network SouthEast boundaries). A variable denoting London outer-suburban services as a separate form of suburban service was not significant.

3.3 Other European Insights

On the back of a major review of European wide monetary valuations of time related attributes, Wardman (2013) conducted analysis of a wide range of time multipliers. The mean values for each of four relevant variables are presented in Table 3, split by country although making no distinction by mode due to small sample sizes[8].

Table 3: Reliability Multipliers by Country

	SDE (κ/φ)	**SDL** (μ/φ)	**Mean Late** (α/β)	**RR** (δ/γ)
Denmark	-	-	2.02 : 0.20 : 8	-
Netherlands	0.86 : 0.09 : 32	1.82 : 0.16 : 31	-	0.60 : 0.09 : 9
Norway	0.71 : 0.29 : 4	2.39 : 0.36 : 6	2.48 : 0.49 : 7	0.19 : 0.03 : 6
Sweden	0.76 : 0.19 : 5	1.70 : 0.45 : 5	4.07 : 0.71 : 5	0.58 : 0.19 : 3
Spain	0.48 : 0.0 : 1	2.10 : 0.23 : 2	-	0.98 : 0.0 : 1
UK	1.20 : 0.46 : 4	2.20 : 0.25 : 15	4.78 : 1.04 : 7	1.22 : 0.14 : 26
All Other	0.53 : 0.07 : 8	1.13 : 0.10 : 7	-	-
ALL	0.81 : 0.07 : 54	1.70 : 0.11 : 66	3.24 : 0.39 : 27	0.91 : 0.10 : 45

Note: The terms are the mean multiplier, its standard error and the number of observations.

Note that, contrary to expectation, the multiplier for late time exceeds that for SDL, generally and overall by a considerable margin. This could well be because the explicit presentation of late time, in terms of being X minutes late one in Y journeys, induces more strategic bias, although late time is also present in the multiple journeys that underpin SDL estimation. Alternatively, it might be argued that the method of presenting travel times in SP exercises intended to estimate SDL are more complex and that

[8] There are slightly fewer observations here for late time than in Table 1 since this review of European evidence focussed on studies with monetary valuations and hence studies providing time multipliers but not money values were not covered. Details of the studies covered here are available from the authors on request.

respondents do not perceive the full extent of the delay involved. We feel that a likely cause of at least some of the discrepancy is that studies valuing late time have estimated mean lateness on the basis that the (Y-1)/Y journeys are all on time. However, if respondents perceived there to be a degree of lateness with these journeys, then its value may have been overestimated. The mean value of late time in Table 3 of 3.24 is a little larger than the railways in Britain currently use. However, removing the UK evidence yields a w_{AML} of 2.7, which is slightly lower than the PDFH recommended value of 3.

With reference to earlier discussion in section 2.1, another strand of research that we can exploit is theoretical work examining the relationship between the mean-variance and scheduling approaches (Bates, 2001; Fosgerau and Karlström, 2009). More specifically, Bates et al. demonstrated the following relationship between the reliability ratio (RR) and the parameters arising from the scheduling approach:

$$RR = \frac{\kappa}{\varphi} \ln\left(1 + \frac{\mu}{\kappa}\right) \tag{8}$$

Taking φ to be one, our results for all the observations have average values of 0.81 for κ and 1.70 for μ. This implies a RR of 0.92, almost exactly the figure reported in Table 3 for the estimated RR. Restricting the comparison to the UK evidence, we find an implied RR of 1.25 compared to an estimated value of 1.22.

Finally, a meta-model was estimated to 1389 European-wide time multiplier observations, including the 202 relating to reliability[9]. The estimated model implied the multipliers set out in Table 4. There was no variation found by mode but the late time multiplier varied by purpose. We do not find it surprising that reliability is relatively less important as distance increases, since longer distance journeys are expected to be less reliable and hence late arrival is less unacceptable. However, this contrasts with the findings of our regression of UK evidence. In addition, the late time multiplier is found to be somewhat larger for

[9] The attributes to which the other multipliers related were walking and waiting time, time spent in congested traffic and searching for a parking space, service headway and departure time shifts.

'Other' trips, which is perhaps surprising, whereas non-commute trips had relatively low values for the UK evidence.

Table 4: Multipliers Implied by European-wide Meta-analysis

Kilometres	**5**	**25**	**100**	**250**
Mean late				
Commute	3.59	2.82	2.30	2.00
EB	3.59	2.82	2.30	2.00
Other	6.71	5.28	4.30	3.75
RR				
All	0.80	0.63	0.51	0.45
SDE				
All	1.02	0.80	0.65	0.57
SDL				
All	2.17	1.71	1.39	1.21

4. DIRECT ELASTICITY EVIDENCE

In the past decade, a body of evidence has emerged of reliability elasticities obtained from direct demand models. These models entail the extension of the forecasting equation (5), such that:

$$\frac{V_{new}}{V_{base}} = \left(\frac{GJT_{new}}{GJT_{base}}\right)^{\eta_{GJT}} \cdot \left(\frac{AML_{new}}{AML_{base}}\right)^{\eta_{AML}} \tag{9}$$

As far as we are aware, directly estimated elasticities are unique to the railway industry in Great Britain.

4.1 Review of UK Evidence

Whilst there is some evidence from choice modelling, such as mode choice models which include measures of reliability and hence can estimate the effects of changes in reliability on market share (TPA, 1992; Steer Davies Gleave, 1994; Burge et al., 2010), they are reliant on SP data. For some of the reasons listed in section 2.5, which can cause a divergence between directly estimated and deduced elasticities, our focus here is on six studies that provide unique, although as we shall see imperfect, insights into this challenging area of how reliability impacts on rail demand.

The PPM statistic, indicating the proportion of arrivals at the destination station within five (for short distance) or ten (for long distance) minutes, and available at train operating company level, is the measure of reliability used in four of the six studies, with AML used in OXERA (2005), and both used in Batley et al. (2011). The attraction of PPM for time series models is that it is readily available over many years. However, our previous discussion of values and PDFH's forecasting framework are both based around mean lateness. We therefore need to convert the PPM elasticities into equivalent AML elasticities to enable useful comparisons to be made with the elasticities implied by equation (6). The relationship between the AML elasticity (η_{AML}) and the PPM elasticity (η_{PPM}) is[10]:

$$\eta_{AML} = \eta_{PPM} \frac{\partial PPM}{\partial AML} \frac{AML}{PPM} \tag{10}$$

Thus we can convert the PPM elasticity into the AML elasticity by multiplying by the elasticity of PPM to AML. We obtained the latter by regressing the logarithm of PPM on the logarithm of AML using data supplied by Network Rail. This data represented PPM and AML at national level for 26 periods between 2008/11 and 2010/10. The adjusted R^2 goodness of fit was 0.85. The intercept term was 0.016 with a t-ratio of 1.7 whilst the elasticity term was -0.11 with a very high t-ratio of 12.0. Thus multiplying the PPM elasticities by -0.11 converts them to AML elasticities.

This conversion is done in Table 5; where a study reports η_{PPM} then we also provide the η_{AML} equivalent. Note that since the elasticity evidence is drawn from analysis of ticket sales data, it is not possible to segment by journey purpose, with the exception that season tickets reflect commuting trips. Distance, ticket type and flow type are the categorisations possible. Table 5 reports significant elasticity estimates by study, flow type and ticket type, and also indicates where the elasticity estimate was not significant and hence not reported. The source of the evidence is given along with the periodicity and

---

10
$$\eta_{AML} = \frac{\partial V}{\partial AML}\frac{AML}{V} = \frac{\partial V}{\partial PPM}\frac{\partial PPM}{\partial AML}\frac{AML}{V} = \frac{\partial V}{\partial PPM}\frac{\partial PPM}{\partial AML}\frac{AML}{V}\frac{PPM}{PPM}$$
$$= \frac{\partial V}{\partial PPM}\frac{PPM}{V}\frac{\partial PPM}{\partial AML}\frac{AML}{PPM}$$

length of the time series data used in estimation and the number of observations per model. The MVA (2009) and ARUP/Oxera (2010) studies used annual data, the MVA (2008) study was based on quarterly data and the remaining three studies used four-weekly data. These time periods define the short run elasticities, but it is the long run elasticities that we should concentrate upon here.

Table 5: Summary Elasticity Table

	Ticket Type (Observations)	**Source**	**PPM**		**AML**	
			SR Elasticity	**LR Elasticity**	**SR Elasticity**	**LR Elasticity**
ARUP/Oxera (2010) Annual data 1990-2008						
LSEE to LSEE	Non Season (71894)	Web	0.43	1.14	-0.05	-0.13
	Seasons (64996)		-	-	-	-
LSEE to Non London Core Cities	Reduced (10982)		0.22	1.26	-0.02	-0.14
	Full (8064)		-	-	-	-
	Seasons (8064)		-	-	-	-
LSEE to Other	Reduced (86310)		-	-	-	-
	Full (14785)		0.62	0.79	-0.07	-0.09
	Seasons (1197)		1.61	2.01	-0.18	-0.22
Non London Core Cities to LSEE	Reduced (6868)		0.31	0.38	-0.03	-0.04
	Full (7269)		-	-	-	-
	Seasons (1297)		-	-4.51	-	0.50
Non London Core Cities to Non London Core Cities	Reduced (2319)		-	-	-	-
	Full (2114)		0.46	0.82	-0.05	-0.09
	Seasons (659)		-	-	-	-
Non London Core Cities to Other	Reduced (16336)		0.25	1.15	-0.03	-0.13
	Full (19106)		0.39	1.22	-0.04	-0.13
	Seasons (6063)		-	-	-	-
Other to LSEE	Reduced (18244)		0.80	1.67	-0.09	-0.18
	Full (13919)		-	-	-	-
	Seasons (4084)		-	-	-	-
Other to Non London Core Cities	Reduced (22387)		0.53	1.53	-0.06	-0.17
	Full (22833)		0.34	1.32	-0.04	-0.15
	Seasons (6022)		0.33	1.24	-0.04	-0.14
Other to Other	Reduced (40938)		-0.09	0.94	0.01	-0.10
	Full (37735)		0.24	1.30	-0.03	-0.14
	Seasons (10590)		-0.06	2.04	0.01	-0.22
Airports	Reduced (8300)		-	-	-	-
	Full (86310)		0.71	0.87	-0.08	-0.10
MVA (2008) Quarterly Data 1980-2007						
London and South East	Seasons (105)	Web	-	-	-	-
	Non Seasons (105)		0.29	0.41	-0.03	-0.05
Regional	Seasons (105)		0.38	0.41	-0.04	-0.05
	Non Seasons (105)		0.25	0.49	-0.03	-0.05
Intercity	Seasons (105)		-	-	-	-
	Non Seasons (105)		0.36	0.58	-0.04	-0.06
MVA (2009) Annual data 1991-2007						
Regional	Seasons (1927)	Unpublished	-	0.56[1]	-	-0.06
	Non Season ≤20miles (695)		-	0.19	-	-0.02
	Non Seasons >20miles (1308)		-	-	-	-
Steer Davies Gleave (2003) 4 weekly data 1995-2001						
Intercity to London	Non Seasons (2441)	Unpublished	0.36	0.39	-0.04	-0.04
Intercity to London	Non Seasons (2441)		-	-	-0.23	-0.25
Intercity from London	Non Seasons (2424)		0.29	0.30	-0.03	-0.03

	Ticket Type (Observations)	Source	PPM		AML	
			SR Elasticity	LR Elasticity	SR Elasticity	LR Elasticity
South East to London	Non Seasons (1522)		0.24	0.28	-0.03	-0.03
South East to London	Seasons (1522)		0.08	0.09	-0.01	-0.01
Non London Inter Urban	Non Seasons (3817)		0.20	0.20	-0.02	-0.02
Batley et al. (2011)[2] 4 weekly data 2002-2007						
Intercity London	Seasons (12474)	Journal	-	-	-0.02	-0.04
	Full (14364)		-	-	-0.01	-0.04
	Reduced (12159)		-	-	-0.01	-0.05
Oxera (2005)[3] 4 weekly data 1995-1999						
Intercity London	All (5140)	Conference	-	-	-0.28	-0.63
London and South East	All (1988)		-	-	-0.06	-0.06
Non London Short	All (1296)		-	-	-0.07	-0.16
Non London Inter Urban	All (24627)		-	-	-0.22	-0.35
Airports	All (3050)		-	-	-0.39	-0.46

Notes: [1] Uses midpoint of the 0.48 to 0.65 range provided. [2] This study also estimated PPM elasticities. These tended to be insignificant but at a coarser level of detail. [3] The long run elasticities here are for three years. LSEE denotes the London and South East and Eastern region. 'Full' denotes undiscounted tickets that can be used to travel at any time. 'Reduced' are lower fare tickets but with travel restrictions attached.

It is interesting to note that the AML elasticities rarely exceed section 2.4's 'illustrative' maximum value of -0.18, but they do often fall short of its illustrative minimum of -0.11. We can also note in Table 5 that the studies that used AML always returned significant coefficient estimates in stark contrast to the results for PPM. This is not necessarily because AML figures were specified at a greater level of detail (i.e. at the flow level), although in one case they were (Batley et al. 2011), but more likely because PPM is a cruder measure of reliability. Another feature to note is that for the season ticket market into London, which is one of the principal sources of rail revenue in Britain, three of the four observations are insignificant and the other is -0.01 (in AML terms). This might reflect the lack of alternatives to rail in this market.

The one exception in terms of detail is that of Batley et al. (2011) which, as far as we are aware, is the only study whose explicit purpose was to determine the impact of reliability on rail demand, and it also employed both PPM and AML measures. Furthermore, this study employed AML data at service group level[11], more detailed than the TOC-level data which had been used in previous studies. In addition, the emphasis in selecting flows for this study was to provide as large a range as possible of reliability

[11] A service group is made up of several station-to-station flows. A TOC will typically consist of several service groups.

changes, mainly driven by changes in timetables to support improved performance. The four-weekly data covered the period 2002 to 2007 and 496 flows, giving rise to some 14,000 observations in the case of non-season, and 12,000 in the case of season. It is worth here replicating the results given the significance of this study to the estimation of reliability elasticities. The elasticities are reported in Table 6 for full fare tickets, reduced fare tickets and season tickets, and for reliability represented as AML and PPM, all estimated by an Autoregressive Distributed Lag (ADL) model.

Table 6: Directly Estimated Reliability Elasticities (Batley et al., 2011)

	Full		**Reduced**		**Seasons**	
	AML	PPM	AML	PPM	AML	PPM
SR	-0.01 (2.5)	0.05 (2.1)	-0.01 (2.0)	0.01 (0.4)	-0.02 (2.5)	0.02 (0.5)
LR	-0.04 (2.2)	0.19 (2.2)	-0.05 (2.8)	0.02 (0.2)	-0.04 (2.9)	0.03 (0.5)

Note: t-ratios in brackets.

Two things are noticeable about the PPM elasticities. First, only the full fare model recovers significant estimates. Second, these elasticities are lower than almost all of the other significant PPM elasticities reported in Table 5. These are challenging findings given that the explicit purpose of this study was to estimate reliability elasticities. When we turn to this study's AML elasticities, all are significant. This is presumably because the AML data was more detailed and is a more refined measure of reliability than PPM. Nonetheless, these elasticities are low, certainly much lower than the other two directly-estimated AML elasticities for Intercity London flows reported in Table 5.

Table 5 reproduces outputs from all models reported in studies that have provided reliability elasticity estimates. A noticeable feature is that in numerous cases no reliability elasticity is reported on the grounds that it was not statistically significant. Summary measures based solely on the statistically significant evidence can be expected to lead to inflated mean values to the extent that insignificant elasticities, even though not necessarily zero in reality, will generally be lower than significant elasticities.

Insignificant reliability elasticities might have occurred for reasons other than the elasticity being inherently small. Limited data sets, noise in the data and correlations with other variables, or simply little variation in late time over the period of analysis, might all be contributory factors[12]. The use of relatively crude TOC-level PPM figures will not help with precision, and for most studies the examination of reliability was incidental and hence great effort would not have been expended in exploring this issue in detail. Nonetheless, it would be prudent to allow for these insignificant elasticities being lower rather than treating them as implicitly the same as significant elasticities.

Table 7 provides summary statistics for the AML elasticities that we have assembled and deduced. There are 36 observations of significant and correct sign elasticities, increasing to 51 observations if we treat as zero those cases where the reliability coefficient was insignificant or wrong sign. As expected, the mean for All Observations is lower than that for Significant Observations, to the extent of 30%. We identified in subsequent regression analysis an outlier observation, the -0.63 for Inter City London flows in Oxera (2005), and its removal reduces the mean elasticities. As part of a consultation process[13], Network Rail raised concerns about the elasticity figures in the ARUP/Oxera (2010) study. We therefore tested removing these elasticities and it can be seen that that this made little difference.

Table 7: Summary Statistics for Long Run Elasticity Evidence

	Significant Observations	All Observations
No omissions	-0.13 : 0.02 : 36	-0.09 : 0.02 : 51
Omit -0.63 outlier	-0.11 : 0.02 : 35	-0.08 : 0.01 : 50
Omit ARUP/Oxera (2010)	-0.12 : 0.04 : 20	-0.11 : 0.03 : 23
Omit ARUP/Oxera (2010) and -0.63	-0.10 : 0.03 : 19	-0.08 : 0.03 : 22

Note: Figures are mean elasticity, standard error of the mean and number of observations.

Looking at the values in a little more detail, to the extent that our relatively small sample allows, and removing the -0.63 outlier, we see from Table 8 that the elasticities in the South East of England, which

[12] Indeed, this is an issue that is widely recognised to affect GJT elasticity estimates. Even though GJT is not a small elasticity, it is not uncommon to find that it is insignificant because of little variation in service quality across the flows and time periods of interest.

[13] This review was undertaken as part of the 2013 PDFH update (PDFH5.1)

cover shorter distance trips into London, are by some margin the smallest and those for Regional are the largest. This is presumably because rail is much more dominant in the South East, in many instances with few real alternatives to rail, whereupon elasticities will be low. On regional flows, rail market share is relatively low and hence we might expect its elasticities, all else equal, to be large. There does not though seem to be a distance effect but, as expected, airport flows have the largest elasticities.

Turning to ticket types, the elasticities for the All Tickets category stand out and these were all obtained from the OXERA (2005) study. For Significant Observations, we note that the Non Seasons category has the lowest elasticity, but then again, Full and Reduced are also non season tickets but yield similar elasticities to the Seasons category. For All Observations, the evidence is that Seasons has the lowest elasticity, but then there is little variation apart from the All Tickets category.

Table 8: Elasticity Evidence by Market Segment and Ticket Type

LR Elasticity	**Significant Observations**	**All Observations**
All	-0.11 : 0.02 : 35	-0.08 : 0.01 : 50
South East	-0.06 : 0.02 : 5	-0.04 : 0.02 : 7
Regional Urban	-0.11 : 0.03 : 6	-0.11 : 0.03 : 6
Regional Interurban	-0.14 : 0.06 : 5	-0.08 : 0.04 : 9
Regional All Flows	-0.12 : 0.02 : 5	-0.12 : 0.02 : 5
Intercity	-0.10 : 0.02 : 12	-0.06 : 0.02 : 20
Airport	-0.28 : 0.18 : 2	-0.19 : 0.14 : 3
Seasons	-0.11 : 0.03 : 7	-0.05 : 0.02 : 15
Full	-0.11 : 0.01 : 7	-0.07 : 0.02 : 10
Reduced	-0.12 : 0.02 : 7	-0.08 : 0.02 : 10
Non Seasons	-0.07 : 0.02 : 10	-0.07 : 0.02 : 11
All Tickets	-0.26 : 0.09 : 4	-0.26 : 0.09 : 4

If we just look at the AML elasticities, thereby avoiding the conversion from PPM to AML, **and** if we remove the -0.63 figure and the large airport figure of -0.46, the average over the remaining six directly estimated AML elasticities is -0.15[14]. This is, interestingly, midway between the illustrative maximum and minimum values of section 2.4.

4.2 Explaining the Elasticity Evidence

We have also used a regression approach to try and explain the quite diverse set of elasticities across studies, just as we did for the late time multipliers. The variables we examined in the regression were:

- Flow types of to/from London, within the South East, non-London, mix of each, and airport flows
- Inter-urban or suburban travel or a mix of the two
- Ticket type
- Whether the AML elasticity was directly estimated or deduced from a PPM elasticity.

Given that all studies were comparatively recent, we did not consider study age. Estimates from linear-additive models of the long run AML elasticities for Significant Observations and All Observations are reported in Table 9, again showing the effect of removing the -0.63 outlier observation. Given the limited amount of data and its diverse nature, coefficients that returned a t-ratio in excess of one were retained. The goodness of fit achieved is low, but that is not surprising given the amount of variation in the elasticities and the inclusion of a large proportion of zero elasticities in the model for All Observations.

Airport flows are, unsurprisingly, found to have larger elasticities, and in both models the effect is significant at the 5% level. Where the elasticities have been deduced, they are smaller for Significant Observations and slightly larger for All Observations. The best-fitting model for Significant Observations showed the elasticities to be smaller in the South East, an effect also apparent in Table 8, presumably

[14] After the review was conducted, we became aware of a confidential study of short and long distance non-season flows into London where the AML elasticity was -0.18 with a 95% confidence interval of only ±0.006 and also a shorter distance largely commuter based flow into London where the highly significant AML elasticity was around -0.25 for both seasons and non-season tickets. Nonetheless, we have not uncovered additional evidence relating to season tickets.

because the alternatives to rail are relatively unattractive for this segment. In the model for All Observations, the best-fitting model was that which instead specified a dummy variable for Seasons, also apparent in Table 8, again indicating that rail is in a relatively strong position for this segment.

Table 9: Regression Analysis of Elasticity Evidence

	Significant Observations	**All Observations**
Constant	0.16 (5.2)	0.07 (2.6)
Airport	0.15 (2.4)	0.11 (2.0)
Deduced	-0.07 (1.9)	0.04 (1.4)
South East	-0.06 (1.4)	
Seasons		-0.03 (1.0)
Observations	35	50
Adjusted R^2	0.24	0.09

Note: t-ratios in brackets.

Whilst admittedly not the most robust of regression models, they do at least provide a means of 'making sense' of the evidence and the results are consistent with the tabulations in Table 8. Taking the directly-estimated elasticities as preferred, the Significant Observations and All Observations models imply AML elasticities of:

- -0.16 and -0.07 for non-airport flows outside of the South East and for non-seasons respectively
- -0.10 and -0.04 for non-airport flows within the South East and for season tickets respectively
- -0.31 and -0.18 for airport flows, excluding the South East effect in the former since the evidence did not relate solely to such flows and the seasons effect in the latter since these are largely irrelevant for airport flows

One possible interpretation of the findings is that the 'true' elasticity might be bounded at the upper end by the elasticity from Significant Observations and at the lower end by the elasticity from All Observations, given that non-significant estimates can be expected to be less than significant estimates but possibly greater than zero.

5. SYNTHESIS

Having reviewed the late time valuations and elasticities, we can assess their consistency. We have a number of sources of summary valuations and elasticities which complicates matters a little. Table 10 reports actual and implied elasticities for the eleven key market segments used in PDFH.

The implied late time elasticities of equation (7) require information on the GJT elasticity, the levels of mean lateness and GJT, and the value of the lateness multiplier. The GJT elasticities used in deducing late time elasticities are taken from the most extensive meta-analysis of such evidence (Wardman, 2012). The meta-model implied GJT elasticities that varied quite strongly with distance and slightly between season and non-season tickets, but did not vary across flow types. Hence Table 10 specifies some representative distances to derive GJT elasticities, which are explicitly long run to be comparable with the directly-estimated elasticities, and these also serve to guide the levels of GJT. The mean lateness figures are based on official statistics. As for the lateness multipliers (w_{AML}), we have used three sets of figures:

- PDFH: we use the recommended multipliers at the time of this research (PDFH5).
- UK: we use the multipliers implied by Model I of Table 2 covering UK evidence.
- EUR: we used the multipliers from the meta-model of European evidence reported in Table 4.

For each set of multipliers, we derived the implied AML elasticity (η_{AML}). To facilitate comparison, the direct elasticities obtained from the models in Table 9 for Significant Observations and All Observations are reported ($\eta_{AML,Sig}$ and $\eta_{AML,All}$), along with the mean of the two ($\eta_{AML,Mean}$), as well as the implied w_{AML} that best fitted the direct elasticities.

The PDFH implied elasticities for commuting are always greater than the direct estimates for Significant Observations $\eta_{AML,Sig}$, although the divergence is small for the Non London flows. This would remain so for the long distance flows, even if the mean lateness was somewhat lower. The differences become very large compared to the mean of the direct estimates on Significant Observations and All Observations $\eta_{AML,Mean}$, in what is a segment that accounts for a large proportion of rail revenue. For non-commuting, the discrepancies between the implied PDFH and Significant Observations elasticities are relatively minor, except for airport flows, but with the exception of the Non London flows the discrepancies become quite appreciable when compared to the mean of the Significant Observations and All Observations estimates.

Turning to the implied elasticities based on our UK review evidence, these are often larger than those implied by PDFH, and some exceed the directly estimated elasticities (substantially in some instances). And where the UK elasticities are less than those from PDFH, it is actually where the latter corresponds well with the direct evidence. Moreover, the review-based evidence seems to be inferior for the purpose of forecasting reliability effects than using the rather crude w_{AML} of 3 throughout. On this basis, we might conclude that PDFH's current recommendations are (relatively speaking) defensible.

With regard to the European evidence, w_{AML} tends to vary around the recommended PDFH value of 3, although it does embody a declining distance effect, and thus provides a distinct contrast to the increasing distance effect in PDFH. Its performance in a forecasting context relative to the other weights is mixed. For Non London flows, it provides the closest elasticities to the direct estimates on Significant Observations and for the Direct Mean it performs as well as PDFH. For commuting to and in London, it performs the best but even then some large differences are apparent, whilst for Non-Commuting it performs the worst.

Excluding airport flows, where the implicit elasticities all seem too low, the elasticities implied by the PDFH forecasting framework tend to exceed the direct estimates on Significant Observations, with some large discrepancies particularly when using w_{AML} from UK evidence. When we compare against the mean of direct estimates on All and Significant Observations, which can reasonably be taken as our best estimate, the implied elasticities tend to be much larger and cannot be regarded as consistent with directly-estimated elasticities.

Finally, and exploiting equation (7), we can examine the w_{AML} that is implied by η_{GJT}, η_{AML}, GJT and the level of mean lateness, again considering estimates based on Significant and All Observations, as well as the mean of the two. As regards All Observations, the implied $w_{AML,All}$ are generally not credible, with several less than zero and all but the airport elasticity less than what the vast majority of evidence would suggest. This is not a surprising finding, given that insignificant directly-estimated elasticities will not be zero. The implied $w_{AML,Sig}$ for Significant Observations and for the mean of All Observations and Significant Observations $w_{AML,Mean}$, despite being somewhat different from each other, do generally seem plausible. The $w_{AML,Sig}$ implied by Significant Observations tends to be less than that directly estimated. However, the most sensible basis for comparison is the $w_{AML,Mean}$ implied by the mean of All Observations and Significant Observations, which tend to be lower (by some considerable margin) than the directly-estimated w_{AML}.

Table 10: Implied and Directly Estimated Elasticities

Segment	Purpose	Miles	η_{GJT}	GJT	Mean Late	w_{AML} [15]	η_{AML}	w_{AML}	η_{AML}	w_{AML} [16]	η_{AML}	$\eta_{AML,Sig}$	$\eta_{AML,All}$ [17]	$\eta_{AML,Mean}$	$w_{AML,Sig}$	$w_{AML,All}$	$w_{AML,Mean}$
						PDFH		**UK**		**EUR**		**Direct**			**Implied**		
Long Distance	Commute	50	-1.44	75	5.2	3.0	-0.30	5.08	-0.51	2.38	-0.24	-0.16	-0.04	-0.100	1.60	0.40	1.00
to/from London	Non Commute	150	-1.56	120	5.2	3.0	-0.20	3.42	-0.23	3.17	-0.21	-0.16	-0.07	-0.115	2.37	1.04	1.71
South East	Commute	30	-1.29	60	2.6	3.0	-0.17	3.92	-0.22	2.56	-0.14	-0.10	-0.04	-0.070	1.79	0.72	1.26
to/from London	Non Commute	30	-1.13	60	2.2	3.0	-0.12	2.26	-0.09	4.03	-0.17	-0.10	-0.07	-0.085	2.41	1.69	2.05
London	Commute	10	-1.04	30	1.6	3.0	-0.17	3.92	-0.22	3.02	-0.17	-0.10	-0.04	-0.070	1.80	0.72	1.26
TravelCard Area[18]	Non Commute	10	-0.91	30	1.2	3.0	-0.11	2.26	-0.08	4.76	-0.17	-0.10	-0.07	-0.085	2.75	1.92	2.33
Non London > 20	Commute	50	-1.44	60	2.7	3.0	-0.19	5.08	-0.33	2.38	-0.15	-0.16	-0.04	-0.100	2.47	0.62	1.55
miles	Non Commute	75	-1.36	90	2.7	3.0	-0.12	3.42	-0.14	3.53	-0.14	-0.16	-0.07	-0.115	3.92	1.72	2.82
Non London < 20	Commute	10	-1.04	40	2.3	3.0	-0.18	3.92	-0.23	3.02	-0.18	-0.16	-0.04	-0.100	2.68	0.67	1.68
miles	Non Commute	10	-0.91	40	1.8	3.0	-0.12	2.26	-0.09	4.76	-0.19	-0.16	-0.07	-0.115	3.91	1.71	2.81
Airport	Non Commute	50	-1.26	75	2.0	6.5	-0.22	4.90	-0.16	3.75	-0.13	-0.31	-0.18	-0.245	9.23	5.36	7.29

[15] We have not here used the PDFH sensitivities referred to in section 2.3 except for airports.
[16] The weights here are from the meta-model that underpins the results reported in Table 4. For non-commuting, a balance of 33% business and 67% other is taken.
[17] All represents the elasticities implied when the non-significant estimates are treated as zero.
[18] The flows covered by this area are typically referred to as inner-suburban.

Although approximations are introduced by the need to use representative values of GJT and mean lateness, we have to conclude that there are significant discrepancies between the directly-estimated late time elasticities and the elasticities implied from valuations of late time. This might be because, as some would suspect, freely estimated late time valuations have been influenced by strategic bias and protest response in SP experiments. However, we cannot be certain that this is the case, and a more sobering explanation would be to conclude that, whilst rail travellers clearly dislike unreliability (as evidenced by the lateness multiplier), they may be unwilling or unable to reduce their rail travel in response to experiences of poor performance.

6. CONCLUSIONS

This paper provides an extensive review of evidence relating to reliability parameters, drawing together and synthesising the findings from numerous British studies and thereby providing new insights in this challenging area. The review of late time valuations brings a significant amount of new evidence into the public domain whilst the coverage of late time elasticities is particularly original since we are not aware of any such review. These reviews put us in the unique position of being able to compare the elasticities implied by the long established use of late time values alongside directly-estimated elasticities.

The evidence relating to both late time valuations and elasticities is diverse and not easily explained. Regression analysis provides some insights but they are limited, although the sample sizes we are dealing with compared to, say, meta-analyses of time valuations need to be borne in mind.

The late time values emerging from UK evidence are generally large and exceed by some margin current PDFH recommendations. Indeed, we find that the values of late time exceed by some degree the values of SDL, and this should not be the case. This might be because late time is more explicit in the former, making it easier to value, and possibly also more prone to protest response given that it is a contentious issue. However, we have also pointed out it may be that the amount of late time has been understated and hence its valuation is exaggerated. The late time multipliers exhibit variation with only a limited number of factors and the positive distance effect contrasts with the negative effect obtained from the cited European-wide meta-analysis. The latter study provides summary values of late time, the reliability ratio, SDE and SDL. What we find is a remarkably high degree of consistency between the SDE and SDL values and the reliability ratio not only for the UK but for the European evidence as a whole. Given this consistency surrounding SDL values, and that the late time values are larger, this adds to the view that the late time values are exaggerated.

The directly-estimated elasticity evidence is also diverse, and a particular issue is that insignificant coefficient estimates for reliability measures are commonplace. These elasticities are unlikely to be zero in practice and hence this adds a significant element of uncertainty into our review. A further issue is that a number of studies used the PPM measure of reliability instead of the preferable Average Minutes Late. Nonetheless, the evidence suggests that reliability does impact on rail demand and we have recovered limited variation by flow type. Noticeably, the elasticity evidence based on mean lateness is more robust than that based on PPM.

We are in the fortunate and original position of being able to compare significant amounts of evidence relating to directly-estimated and implied late time elasticities. Although there are assumptions and approximations involved in this process, it is perhaps the most important aspect of this paper. We find that the late time multipliers imply elasticities somewhat larger than the directly-estimated elasticities. This is consistent with other aspects of the paper, in particular the proposition that the SP-based late time valuations are too large, but we have also discussed other reasons why the implied elasticities might exceed directly-estimated ones.

Although we have covered a significant amount of material, it points to a need for further work in this area. Firstly, the SP values might be too large, but this needs to be tested by obtaining values from well-defined RP choice contexts offering clear trade-offs between reliability and other variables and essentially with large sample sizes. Secondly, the other reasons why the direct and implied elasticities might differ, generalising to other similar contexts, need to be further examined. Finally, it is clear that much further work is needed on obtaining robust directly-estimated late time elasticities. This should not be treated as a 'side issue', routinely entering some readily available reliability variable alongside other terms in a rail demand model. This review demonstrates that studies which enter the more appropriate mean lateness rather than PPM are more successful in recovering significant effects, as are studies that pay attention to detail in terms of selecting suitable flows and detailed data that support reliable estimates. In due course, suitable measures representing the variability of lateness in addition to mean lateness should be entered into these demand models.

Finally, whilst we have focused almost entirely on British evidence in the rail market, in our understanding this has by far the longest pedigree of explicitly handling reliability in appraisal. We recommend not only that other railway administrations, but also transport appraisal in general, adopt an appropriate approach in this area.

REFERENCES

Abrantes, P.A. L. and Wardman, M. (2011) Meta-Analysis of U.K. Values of Time: An Update. Transportation Research A, 45 (1), pp.1-17.

ARUP/OXERA (2010) What are the Findings from the Econometric Analysis? Findings Report. Prepared for the Department for Transport, Transport Scotland and PDFC

Association of Train Operating Companies (2013) Passenger Demand Forecasting Handbook Version 5.1. London.

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

Bates, J. (2010) Stock-take of Travel Time Variability. Prepared for Department for Transport. http://www.dft.gov.uk/publications/stock-take-of-travel-time-variability

Batley, R., Dargay, J. and Wardman, M. (2008) Revealed Preference Study to Assess Impact of Reliability on Passenger Rail Demand. Deliverable 4: Report of Econometric Analysis. Final report to the Department for Transport.

Batley, R.P., Dargay, J.M. and Wardman, M. (2011) The Impact of Lateness and Reliability on Passenger Rail Demand. Transportation Research E 47, pp.61-72

Batley, R. and Ibáñez, J.N. (2013) Randomness in preference orderings, outcomes and attribute tastes: An application to journey time risk. Journal of Choice Modelling, 5 (3), pp157-175.

Benwell, M. and Black, I.G. (1984) User Valuation of Reliability in Inter-Urban Travel: Some Problems of Modelling and Measurement. Proceedings of Seminar L, PTRC Summer Annual Meeting, PTRC, London.

Black, I.G. and Towriss, J.G. (1993) Demand Effects of Travel Time Reliability. Prepared for Department for Transport.

Burge, P, Rohr, C. and Kim, C. (2010) Modelling Choice for Long Distance Travel in the UK: An Analysis of Mode Choice. Proceeding of European Transport Conference.

Carrion, C. and Levinson, D. (2012) Value of Travel Time Reliability: A Review of Current Evidence. Transportation Research A 46 (4), pp. 720-741.

Department for Transport and Office of Rail Regulation (2012) Realising the Potential of GB Rail – Detailed Report: Final Independent Report of the Rail Value for Money Study.

Faber Maunsell (2003) Outward Facing Research – Journey Time Reliability Stated Preference Survey. Final Report to Highways Agency.

Faber Maunsell and Mott MacDonald (2007) Rail Overcrowding, Reliability and Frequency Demand Impacts Project. Final Report to Centro.

Fosgerau, M. and Karlström, A. (2009) The Value of Reliability. Transportation Research Part B, 44 (1), 38-49.

Hollander, Y. (2006) Direct versus Indirect Models for the Effects of Unreliability. Transportation Research A, 40, pp.699-711.

Hollander, Y. (2009) What do we really know about travellers' response to unreliability? Paper presented at the International Choice Modelling Conference, Harrogate, UK, 2009.

Jackson, B.W. and Jucker, J.V. (1981) An empirical study of travel time variability and travel choice behaviour. Transportation Science, 16 (4), pp.460-475.

Jones, P., Polak, J. and Bates, J. (1995) The Importance of Punctuality and Reliability: A Review of Evidence and Recommendations for Further Work. Report to Passenger Demand Forecasting Subscription Service.

London Transport (1983) 1983 Rail Priority Survey.

Lu, H., Fowkes, A.S. and Wardman, M. (2008) Amending the Incentive for Strategic Bias in Stated Preference Studies: A Case Study in the Users' Valuation of Rolling Stock. NEED PUBLICATION

MVA (2009) Regional Flows: Regional Rail Demand Elasticities. Report to Passenger Demand Forecasting Council.

MVA (2008) Econometric Analysis of Long Time Series Rail Passenger Demand Aggregates. Prepared for Department for Transport

MVA (2000) Valuation of Reliability and Punctuality Improvements on Rail Services. Final Report to Shadow Strategic Rail Authority

MVA (1991) Market Research for ORCATS Model. Prepared for InterCity, British Railways Board.

MVA (1989) Network SouthEast Quality of Service Phase III Final Report. Prepared for Network SouthEast, British Railways Board.

MVA, ITS Leeds University and TSU Oxford University (1987) Value of Travel Time Savings. Policy Journals, Newbury, Oxford.

Neumann, J. Von and Morgenstern, O. (1947) Theory of Games and Economic Behavior (Second Edition). Princeton University Press, Princeton.

Noland, R. and Small, K. (1995) Journey time uncertainty, departure time choice and the cost of the morning commute. Working paper UCI-ITS-WP-95-1, Institute of Transportation Studies, University of California, Irvine.

OXERA (2005) How Do Rail Passengers Respond to Change? Prepared for Passenger Demand Forecasting Council.

Passenger Focus (2012) Passenger Focus Response to the Office of Rail Regulation Consultation “Periodic Review 2013: Consultation on Incentives”.

Steer Davies Gleave (1994) M1/A1(M) Corridor Study. Prepared for Department of Transport.

Steer Davies Gleave (2003) The Effect of Road Congestion on Rail Demand. Report to Passenger Demand Forecasting Council

Steer Davies Gleave (2008) New Generation Transport in Leeds: SP Study. Metro and Leeds City Council.

TPA (1992) Trans-Pennine Rail Strategy Study: Model Calibration and Validation Report. Prepared for GMPTE, Merseytravel, SYPTE, WYPTE, Cheshire CC, Derbyshire CC, Humberside CC, Lancashire CC and Peak Park.

Tseng, Y.Y. (2008) Valuation of Travel Time Reliability in Passenger Transport. Research Series Tinbergen Institute, Vrije Universiteit, Amsterdam

Wardman, M. (2014) Price Elasticities of Surface Travel Demand: A Meta-Analysis of U.K. Evidence. Journal of Transport Economics and Policy 48 …. To be confirmed before proofs.

Wardman, M. (2013) Value of Time Multipliers: A Review and Meta-Analysis of European-Wide Evidence. Presented at Annual Conference of the U.S. Transportation Research Board

Wardman, M. (2012) Review and Meta-Analysis of U.K. Time Elasticities of Travel Demand. Transportation 39 pp.465-490.

Vickrey, W.S. (1969) Congestion theory and transport investment. *American Economic Review*, 59 1059 (2), pp.251-260.

WS Atkins (1997) Bus Reliability Study: Stated Preference Research. Prepared for Department.