



# The demand impacts of train punctuality in great britain: systematic review, meta-analysis and some new econometric insights

Mark Wardman<sup>1</sup> · Richard Batley<sup>1</sup>

© The Author(s) 2021

## Abstract

This paper updates and extends the systematic review and meta-analysis of Wardman and Batley (Transportation 41:1041–1069, 2014), which hitherto was the most comprehensive study of the impacts of punctuality on passenger rail demand in the literature. Whereas the 2014 paper covered 51 elasticities from 6 studies in Great Britain published between 2003 and 2011, this updated paper adds 11 subsequent British studies yielding a further 201 observations. The meta-model recovers a range of significant effects, relating to whether the elasticity was short versus long run, flow type and distance, season versus nonseason tickets, the relevant measure of lateness, and whether the purpose of the study was specifically the estimation of late time elasticities. Allowance was also made for study quality-related issues. The data indicated that, despite dynamic models being commonplace, there is some uncertainty as to how long the long run is. Alongside the meta-model, the paper also reports new econometric evidence that addresses some gaps in existing evidence and knowledge, especially in relation to functional form and non-linearity of effects. Findings from both strands of analysis would seem to suggest that rail industry guidance has tended to overstate the demand impacts of punctuality.

**Keywords** Demand elasticity · Late time · Reliability · Meta-analysis · Review · Rail demand

## Introduction

### Context

The literature contains a number of major reviews and meta-analyses addressing a wide range of parameters used in transport planning and appraisal. In the specific context of

---

✉ Mark Wardman  
mwardman@outlook.com

Richard Batley  
r.p.batley@its.leeds.ac.uk

<sup>1</sup> Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

travel time variability, most empirical evidence relates to travellers' valuations of late time and hence the reviews have almost invariably focussed on this. However, the effects on travel behaviour are very important, and Wardman and Batley (2014) discuss its recent significance in the context of the railway industry in Great Britain which has pioneered the investigation of how changes in train punctuality impact upon demand. This has been facilitated by the existence of enormous amounts of recorded station-to-station ticket sales over numerous years, which has long supported econometric analysis of a wide range of impacts on rail demand that in the 2000s was extended to late time as a result of the emergence of temporal data regarding train running performance.

Wardman and Batley (2014) conducted a systematic review and meta-analysis of late time elasticities for rail travel estimated in Great Britain that existed up to 2011. There is now a much more significant body of evidence and this, in part, motivates the present paper.

We also report fresh econometric insights based on the analysis of large datasets available to us. The motivation here is that despite a number of significant recent studies, there is scope to add to the understanding of how late time impacts on rail demand. An underlying theme of both strands of research is addressing widespread concerns that the parameters used to forecast late time variations in the rail industry in Great Britain will exaggerate demand sensitivity.

## Objectives

The aims of this paper are to:

- update and extend Wardman and Batley (2014) which, at the time of publication, formed the most comprehensive review and meta-analysis of late time elasticities;
- summarise recent research and practical recommendations regarding late time elasticities;
- report new econometric analysis that addresses gaps in existing evidence;
- provide recommendations for further research in the area.

The structure of the paper follows these aims. “[Background](#)” section provides relevant background in terms of the method and parameters used in Great Britain to forecast the demand impacts of changes in late time along with a summary of recent studies and relevant policy issues. “[Meta-analysis](#)” section reports the updated meta-analysis and its implied late time elasticities whilst new econometric insights based upon the analysis of large data sets of rail demand are provided in “[New econometric insights](#)” section. Concluding remarks are contained in “[Conclusions](#)” section.

## Background

### The rail network in great Britain

To provide some context to rail services in Great Britain, there is a well-developed network of nearly 10,000 miles and over 2500 stations connecting almost all significant settlements. The section of the GB network covering London and the South East of England is among the densest and busiest networks in the world, with other intensive parts of the network

centred upon major metropolitan areas outside of the capital. Whilst high speed trains are presently limited to those from London to the Channel Tunnel and beyond, major towns and cities are connected by fast and frequent services using high quality rolling stock. Service frequencies of every 15 min or better are common on suburban services whilst it is rare that inter-urban service frequencies are worse than every hour. Freight services share much of the same infrastructure. Great Britain has the fifth highest level of train use in the world, with passenger numbers more than doubling since 2000 and crowding common particularly on the suburban networks. The dense network, intensive use and high service frequencies, together with sustained demand growth, mean that the railways in Great Britain face significant and to some extent unique challenges in providing punctual train services. There is clear evidence that lateness varies across routes and time periods and by distance, all of which are to be expected. Typical levels of mean lateness are in the range 1½ to 3 min for suburban services and 3 to 5 min for longer distance services. Further details on the rail network in Great Britain, its usage and punctuality metrics are provided in Rail Delivery Group (2017, 2021) and Department for Transport (2019).

### Forecasting changes in late time in the rail industry in Great Britain

The Passenger Demand Forecasting Handbook (PDFH) is unique amongst railway organisations and indeed transport administrations worldwide in that, since 1986, it has provided demand forecasting parameters covering many exogenous and endogenous variables based on a synthesis of best available evidence (Rail Delivery Group 2018). It is regularly updated, with the latest version (v6) released in 2018. Recommendations relating to late time have been included since the first edition. Wardman and Batley (2014) stated that the PDFH, “..... would seem to be one of the earliest, if not the earliest, treatments of reliability in a demand forecasting context”.

Prior to the sixth edition, the demand impacts were inferred by converting late arrival time into equivalent journey time and then applying a time elasticity. This is termed the ‘indirect’ approach, with the volume of rail demand ( $V$ ) between any two stations in the new period relative to the base period forecast as:

$$\frac{V_{new}}{V_{base}} = \left[ 1 + \frac{\varpi (APM_{new} - APM_{base})}{GJT_{base}} \right]^{n_{GJT}} \quad (1)$$

where  $\varpi$  is the multiplier expressing the value of late time in equivalent time units and late time is measured by Average Performance Minutes (APM) composed as:

$$APM = AML + DML \quad (2)$$

AML denotes the average minutes of lateness at a destination station, with early arrivals treated as being ‘on time’.<sup>1</sup> Deemed Minutes Late (DML) covers cancellations and is specified as 1.5 times the service interval. The railway industry in Great Britain widely uses, as did some early econometric models, what is termed the Public Performance Measure (PPM), which indicates the proportion of trains that arrive at their terminating station

<sup>1</sup> Earliness is assumed to hold no positive or negative value for passengers, and thus has no impact on demand. Very occasionally, as discussed in Wardman and Batley (2014), the standard deviation of late time has been considered.

within 5 min of schedule for shorter distance operators and within 10 min for longer distance operators.

Equation 1 also contains Generalised Journey Time (GJT) which expresses the journey time (T), service headway (H) and the number of interchanges (I) between stations in equivalent travel time units as:

$$GJT = T + \gamma H + \phi I \quad (3)$$

where  $\gamma$  and  $\phi$  are the headway and interchange time penalties, whilst  $\eta_{GJT}$  denotes the elasticity of demand with respect to GJT. Estimates of  $\alpha$ ,  $\gamma$  and  $\phi$  are invariably obtained from Stated Preference (SP) experiments and PDFH provides a set of recommended values for these parameters. GJT is composed as a weighted average across the day according to when people are deemed to want to travel and the level of service offered at different times of day.

Equation 1 is essentially an approximation to extending GJT to include late time and an implicit property is that the implied late time elasticity will depend upon the proportion that late time forms of GJT.<sup>2</sup>

PDFH v5.1 issued in 2013 considered the case for migrating from this indirect approach to a 'direct' approach based on changes in lateness defined as APM:

$$\frac{V_{new}}{V_{base}} = \left( \frac{APM_{new}}{APM_{base}} \right)^\eta \quad (4)$$

where  $\eta$  denotes the late time elasticity, typically obtained from econometric models of rail demand. However, having reviewed the conceptual and empirical arguments, v5.1 retained the indirect approach but provided further revisions to  $\alpha$  as a result of comparing the elasticities implied by (1) with the emerging body of directly estimated elasticities.

By 2018 the balance of arguments was judged to have shifted, and the sixth edition of PDFH recommended the direct approach given the following concerns surrounding the indirect approach:

- The variation in implied late time elasticities according to the proportion that late time forms of GJT can be very large but is imposed by the mathematical formulation rather than empirically demonstrated.
- There is a reliance on late time multipliers drawn from SP studies and concern that these can be exaggerated due to strategic response bias.
- The average late time multipliers obtained from SP studies of rail users are not necessarily appropriate at the margin where behavioural change occurs. Directly estimated demand elasticities will better reflect the marginal effects.
- The indirect approach is based on actual measures of reliability which might not align with travellers' perceptions. These misperceptions will be better accommodated in the elasticities estimated by the direct approach.
- The indirect approach does not permit mitigation, such as travellers taking different trains in order to increase the likelihood that they arrive at their destination on time.

The currently recommended PDFH late time elasticities of the direct approach are based upon the Wardman and Batley (2014) review. Subsequently, the railway industry

<sup>2</sup> We return to the issue of extending GJT to include late time in "New econometric insights" section.

commissioned two studies with the specific purpose of investigating the impact of late time on rail demand (OXERA and Winder Phillips 2017<sup>3</sup>; Steer 2019) and these provide some important new insights to which we now turn.

## Key recent insights

OXERA and Winder Phillips (2017) estimated two-way fixed effects models for 63 four weekly periods covering 2012 through to 2017 for 295 London based flows and 14,364 observations.<sup>4</sup> Bespoke late time APM data at flow level was collected. The estimated models distinguished short run effects (here four weeks) from long run effects where the full behavioural response has worked through,<sup>5</sup> and between anytime day, off-peak day and season tickets, and returned highly significant parameter estimates. The estimated long run proportional elasticity parameter, in the form of Eq. 10 below, for season tickets was  $-0.020$ , implying that a 1 min increase in late time would reduce rail demand by 2%. For anytime and off-peak tickets the corresponding figures were  $-0.038$  and  $-0.035$ , which would result in around a 3.6% reduction in demand for an extra minute of late time. The lesser effect in the commuting market is to be expected.

Steer (2019) extended coverage to all the key geographical segments in PDFH. The four weekly data covered 2013 to 2018 and again bespoke flow level APM data was collected. The two-way fixed effects models distinguished short and long run effects and were estimated to very large datasets of 66,139 observations from 1481 flows for anytime tickets, 64,754 observations from 1470 flows for off-peak tickets, and 42,585 observations from 1205 flows for season tickets. Of 21 market segments, made up of seven PDFH related flow types and the aforementioned three ticket types, significant long run elasticities were obtained for only 10 (48%) of them. The two long run proportional elasticity parameters for season tickets were  $-0.016$  and  $-0.018$  indicating that a minute extra of late time would reduce rail demand by around 1.7%. The corresponding figures for eight non-season tickets varied between  $-0.010$  and  $-0.049$ , with an average of  $-0.023$ . These are relatively minor demand impacts, although again with a lower impact on commuter demand as might be expected.<sup>6</sup>

A number of new insights were provided by these two recent studies.

First, given that, in the case of punctuality, proportional elasticities were considered a priori to be more appropriate than the constant elasticities adopted as standard in PDFH for other drivers of demand, a key question is whether they provide statistically superior models. This was tested in Steer (2019), and in the Steer Davies Gleave (2017) revisiting of the OXERA and Winder Phillips (2017) study, and was indeed found to be the case.

---

<sup>3</sup> A peer-review of this study was undertaken and it took the opportunity to explore a number of additional models (Steer Davies Gleave 2017). We make use of this study subsequently, and indeed its results were selected for inclusion in the meta-analysis dataset discussed below.

<sup>4</sup> The first study to specifically address the issue (Batley et al. 2011) had four weekly data covering 228 flows and 11,400 observations for anytime tickets and slightly lower numbers for season and off-peak tickets.

<sup>5</sup> The core models reported in this study and the Steer (2019) and Steer Davies Gleave (2017) reports adopted the two-way fixed-effects approach containing just two explanatory variables other than the fixed effects: the APM term and a lagged dependent variable.

<sup>6</sup> And this does not account for the season ticket elasticities being insignificant in key South East and short distance regional markets.

Second, it might be argued that some more general function, such as Eq. 11 below, would provide a better fit than either specific cases of constant or proportional elasticities. This was found to be so when tested in Steer (2019) and Steer Davies Gleave (2017).

Third, the recent studies all analysed four weekly data and there are obvious attractions in this given that performance can vary considerably across a year, and noticeably they find that the long run is not very long. OXERA and Winder Phillips (2017) found that on their largely London based flows the long run is reached in a little less than a year for anytime and off-peak tickets and about a year and a half for season tickets. Steer (2019) found that the long run is reached in less than a year, with around 6 months not uncommon.

Fourth, a number of specific insights were provided. OXERA and Winder Phillips (2017) found that too little weight is currently attached to cancellations within APM, that as might be expected the late time elasticity was larger where the frequency was lower, but that there was no difference between the impacts of improvements and deteriorations in performance. Steer Davies Gleave (2017) and Steer (2019) found that the late time elasticity falls with the level of GJT, whilst the latter study explored whether the level of APM had an effect but the results were inconclusive. These studies concluded that further research into these and other possible influences on late time elasticities was warranted.

Finally, comparison of the findings of these recent studies with the recommendations of PDFH, based as they are on older and arguably less reliable studies, is of interest. The relevant figures, converted to mean elasticities, are presented in Table 1.

The Steer Davis Gleave (2017) London Travel Card Area (TCA) evidence can be discounted as far too large. Whilst on some routes there will be close competitors, particularly in the form of underground services or indeed other rail routes, ticket sales data for London TCA provides a notoriously unreliable account of station-to-station movement given the widespread use of area wide travel cards. Not only do the Steer (2019) results suggest that the PDFH recommendations are too large, and generally by a considerable amount, they are also somewhat different to the Steer Davies Gleave (2017) results, notably in terms of being unable to recover significant elasticities for key season ticket flows. The latter study, after discounting the London TCA results, provides results for the South East to London market that would at least support current PDFH recommendations although that would not be the case for season tickets on Non-London short distance flows.

In summary, further research is required given the gaps in the evidence particularly in key markets, although the non-significant results could be symptomatic of very low elasticities, and some noticeable differences in the findings of the two recent bespoke studies. Nonetheless, on balance it seems that current PDFH recommendations are too large. Whilst meta-analysis is not a panacea, it can provide useful explanations of the available evidence which goes beyond the bespoke studies.

## The problems of estimation and policy implications

An issue that is particularly relevant to the estimation of late time elasticities is that a priori they are expected to be relatively small, and could be very small, and hence obtaining robust estimates is challenging.<sup>7</sup> This is not helped by the accurate estimation of late time for inclusion in econometric models being far from straightforward and limitations as to how far back inter-temporal data is available. Compounding this

<sup>7</sup> This challenge also applies to the estimation of cross-elasticities in rail demand models. In that context, economic theory has been harnessed to assist in obtaining more robust estimates than can be obtained from the econometric analysis of rail demand data alone (Wardman et al. 2019).

**Table 1** PDFH recommended and recently estimated APM elasticities

Flow	PDFH (Wardman and Batley 2014)		Steer Davies Gleave (2017)		Steer (2019)	
	Season	Non-Season	Season	Non-Season	Season	Non-Season
London TCA	-0.070	-0.085	-0.176	-0.428 (Anytime) -0.243 (Off-Peak)	n.s	n.s
South East to/from London	-0.070	-0.085	-0.108	-0.100 (Anytime) -0.149 (Off-Peak)	n.s	-0.033 (Anytime)
Long Distance London	-0.100	-0.115	n.a	n.a	n.s	-0.035 (Off-Peak)
Non-London ≤20 miles	-0.100	-0.115	-0.041 <sup>a</sup>	n.s	n.s	-0.096 (Anytime) -0.088 (Off-Peak)
Non-London > 20 miles	-0.100	-0.115	n.a	n.a	-0.041	-0.029 (Anytime) -0.050 (Off-Peak)
Airports	n.a	-0.245	n.a	n.a	n.s	n.s

Steer provided APM data to convert the semi-elasticities in Steer Davies Gleave (2017) and Steer (2019) to implied average elasticities

<sup>a</sup>Whilst these flows were not entirely within the 20 mile limit, they are more suburban than inter-urban in nature

is the fact that variations in late time tend to be relatively small, whilst some studies yielding elasticities were not focussed on lateness, but incidentally included it given that a measure of late time was available. All this means that it can be difficult to distinguish between low or essentially zero elasticity and an unreliable estimate.

This is an issue for meta-analysis since assuming insignificant estimates to be zero could lead to different conclusions compared to removing insignificant estimates as unreliable evidence. Against this background, Wardman and Batley (2014) considered three alternative treatments of the meta-data, namely:

- a) omitting estimates which were not significant, thereby effectively delivering an upper bound;
- b) setting estimates which were not significant to zero, effectively delivering a lower bound;
- c) taking a simple average of (a) and (b).

This problem has previously raised its head in the GB policy domain, since PDFH late time elasticities are also a key input to the Schedule 4 (planned disruption) and 8 (unplanned disruption) performance regimes used by the Office of Rail and Road (ORR) to determine financial flows between Network Rail (NR) and Train Operating Companies (TOCs) to compensate for the revenue loss associated with late running trains. In this context, the choice of approach is a matter of considerable contention since, based on the available empirical evidence, the direct approach invariably indicates a smaller impact of performance on revenue than the indirect approach, with implications for the respective Schedule 4/8 payments borne by NR and the TOCs.

In 2018, ORR embarked upon their most recent update to Schedule 4/8. In the course of this update, the question of which methodological approach to employ for forecasting the impacts of late running trains on revenue once again arose as a significant area of debate but with different considerations applying to two distinct sub-markets, as follows.

### **Treatment of commuting flows in London and South-East**

The new late time elasticity evidence from OXERA and Winder Philips (2017) was accepted for use on London and South East (LSE) commuter flows.

### **Treatment of other flows**

Since the OXERA and Winder Philips (2017) evidence was focussed upon LSE commuting, it was decided that the indirect elasticity of demand should be retained for non-LSE flows (i.e. as per PDFH v5.1), and it was agreed that Wardman and Batley's (2014) meta-model of late time multipliers should continue to be used to derive the relevant multipliers. However, a second area of debate was sparked by the alternative treatments of the meta-data in Wardman and Batley (2014). Given disagreement between NR and the TOCs on which treatment should be adopted, the ORR was eventually called upon to make a ruling, and they ruled that treatment a) above from Wardman and Batley (2014) should be adopted.



In what follows, we include insignificant late time elasticities in our review so as to avoid any inherent bias towards larger elasticities.

## Meta-analysis

### Data assembly

The meta-analysis builds upon the dataset assembled by Wardman and Batley (2014) which covered 51 elasticities derived from 6 studies published between 2003 and 2011. First of all, the 6 previous studies were revisited and more evidence extracted which increased the sample to 84. In particular, we identified elasticities based on APM and PPM in Batley et al. (2011) whilst MVA (2008) and Arup and OXERA (2010) yielded a number of additional elasticities that were insignificant or wrong sign to provide a more comprehensive account of these studies.<sup>8</sup>

A further 11 studies were identified, with the most recent reporting in 2020, which in total yielded an additional 201 elasticity observations. Hence the new dataset consists of 17 studies, listed in “Appendix”, and 285 elasticities. Evidence has not been included from studies that used highly aggregate representations of late time, such as Mott McDonald and University of Southampton (2014) and Gudgeon et al. (2015) where national PPM was used. We have, however, included insignificant and wrong sign late time elasticities from the included studies since not to do so would bias the evidence.

As far as multiple elasticity observations per study are concerned in the overall dataset, 4 studies provided up to 5 observations, 5 yielded between 6 and 10, 3 supplied between 11 and 20, 4 provided between 21 and 40, and the remaining study (Steer, 2019) yielded 60 (21%) observations.

The variables upon which information was collected for the purpose of explaining variations in elasticities within and across studies are:

- Whether the demand responsiveness measure is a constant or proportional elasticity.
- The measure of lateness, which can be PPM, AML or APM, and its spatial detail.
- The year and status of the publication.
- Whether the primary purpose of the study was to estimate late time elasticities.
- Flow type and distance.
- Ticket type.
- Where possible, the t ratio of the elasticity estimate and otherwise whether the demand parameter was significant at the 30%, 10%, 5% or 1% level, along with the number of observations of the estimated model.
- Whether the elasticity is explicitly short run or long run, or neither, and the form of model used in estimation.
- Whether the estimation data was four weekly, quarterly or annual along with the years covered.
- The length of the long run and the lag structure.
- The mean levels of late time, GJT and their ratio in the estimation data.

---

<sup>8</sup> We identified 9 cases where an insignificant late time elasticity had been removed from a model that otherwise merited inclusion in our dataset. We flagged these unreported estimates and entered them as zeros. Wrong sign elasticities were recorded as zero.

**Table 2** Key features of meta-data

	Original	Additional	Total
Proportional elasticity	0 (0%)	188 (94%)	188 (66%)
Constant elasticity	84 (100%)	13 (6%)	97 (34%)
PPM	55 (65%)	2 (1%)	57 (20%)
APM	9 (11%)	155 (77%)	164 (58%)
AML	20 (24%)	44 (22%)	64 (22%)
Time series	18 (21%)	0 (0%)	18 (6%)
Panel	66 (79%)	201 (100%)	267 (94%)
Short run	32 (38%)	52 (26%)	84 (30%)
Long run	32 (38%)	52 (26%)	84 (30%)
Static	20 (24%)	97 (48%)	117 (40%)
Four weekly data	45 (54%)	99 (49%)	144 (51%)
Quarterly data	18 (21%)	32 (16%)	50 (17%)
Annual data	21 (25%)	70 (35%)	91 (32%)
Service code	27 (32%)	103 (51%)	130 (46%)
Train operator	21 (25%)	2 (1%)	23 (8%)
Flow	0 (0%)	96 (48%)	96 (34%)
Sector	36 (43%)	0 (0%)	36 (12%)
< 1000 observations	19 (23%)	12 (6%)	31 (11%)
1000–4999 observations	16 (19%)	61 (30%)	77 (27%)
5000–9999 observations	18 (21%)	37 (19%)	55 (19%)
≥ 10,000 observations	31 (37%)	91 (45%)	122 (43%)
Not significant at 30%	15 (18%)	33 (16%)	48 (17%)
Significant at 30%	5 (6%)	14 (7%)	19 (7%)
Significant at 10%	3 (4%)	11 (6%)	14 (5%)
Significant at 5%	24 (28%)	6 (3%)	30 (10%)
Significant at 1%	37 (44%)	137 (68%)	174 (61%)
Explicit late impact study	27 (32%)	96 (48%)	123 (43%)

We here compare at the level of constant elasticities, since the recent proportional elasticity evidence contains mean levels of late time from which constant elasticities can be derived but older studies which yield constant elasticities tend not to provide the information needed to deduce proportional elasticity parameters.<sup>9</sup>

Table 2 provides summary statistics relating to the explanatory variable data, distinguishing between the original Wardman and Batley (2014) data, the newly assembled data and the combined data.

In the original study, all the evidence took the constant elasticity form and the majority was based on the PPM measure. Of the 201 additional observations, only two relate to PPM whilst 94% are of the proportional elasticity form reflecting the industry's movement towards this approach. APM elasticities form a large majority of the additional evidence and 58% of the total dataset with similar proportions each for AML and PPM elasticities.

<sup>9</sup> Inspection of the recent Steer Davies Gleave (2017) and Steer (2019) results found the estimated constant elasticities to be broadly similar to the average elasticities implied by the proportional elasticity models.

**Table 3** Summary GJT and AML figures in meta-data studies

	GJT	AML
All	102 (4.3) [237] {39:227}	3.36 (0.09) [199] {1.93:4.98}
Urban	49 (1.8) [56] {29:61}	2.77 (0.12) [54] {1.49:3.89}
Inter-Urban	136 (5.3) [137] {80:234}	3.89 (0.13) [105] {2.11:5.61}
Urban and Inter-urban <sup>A</sup>	73 (8.0) [44] {41:101}	2.77 (0.12) [40] {1.42:3.53}
London TCA	44 (4.4) [19] {29:72}	2.80 (0.19) [19] {1.31:3.78}
South East	84 (4.2) [55] {51:104}	2.77 (0.14) [52] {1.45:4.07}
Non London Short <sup>B</sup>	51 (3.5) [32] {36:57}	2.56 (0.15) [30] {1.39:4.08}
Non London Long <sup>B</sup>	143 (13.1) [40] {51:251}	3.53 (0.13) [36] {2.58:4.63}
London long	190 (11.6) [28] {91:267}	4.94 (0.30) [26] {2.86:6.67}
Multiple flow types <sup>C</sup>	98 (6.2) [63] {31:130}	3.86 (0.16) [36] {2.42:4.98}

Figures are mean, (standard error), [number of observations] and {10th percentile: 90th percentile}

<sup>A</sup> This is where the evidence does not distinguish between urban and inter-urban trips. <sup>B</sup> Short (long) is up to (more than) 20 miles. <sup>C</sup> This is where the evidence does not cover a specific flow type

All of the additional evidence is drawn from panel data models compared to 79% previously. Elasticities from dynamic models are well represented, with explicitly short and long run elasticities forming 76% of the original and 52% of the additional. Around half of the assembled evidence was estimated to four-weekly demand data,<sup>10</sup> which is not surprising given that late time can vary markedly even in short time periods, with annual data the next most popular and not a great deal of difference between the previous and new datasets. Flow level data is clearly preferable, given that the models are estimated to demand between stations, but it is only recently that such data has been used in studies largely because the extraction and processing of such data is considerably more onerous. The most aggregate data is at sector level<sup>11</sup> and then train operator level, and this is largely associated with PPM data used in early studies and is rarely used in more recent studies. Service code level data is most common and represents late time at the level of a corridor.

The GB railway industry has at its disposal an enviable amount of station-to-station ticket sales data and this is apparent in the number of observations underpinning model estimation, particularly in more recent years. Very few of the models were estimated to fewer than 1000 observations yet not far from a half of the elasticities were estimated on more than 10,000 observations. This should be conducive to the estimation of very precise late time elasticities, and this is apparent in the figures indicating levels of significance despite the estimation of late time elasticities facing greater challenges than for most other explanatory variables as discussed in “[The problems of estimation and policy implications](#)” section.

Studies having the explicit purpose of estimating late time elasticities in the original (Batley et al. 2011) and additional datasets (Steer Davies Gleave 2017; Steer 2019) yield 43% of elasticities overall.

<sup>10</sup> In all cases, the late time variables in the demand models are specified at the same periodicity as the demand data.

<sup>11</sup> The sectors in Great Britain are London and the South East, Inter-City and Regional, with further breakdown into season tickets and non-season tickets.

**Table 4** Summary short run, long run and static elasticities

	Constant	Proportional	APM	AML	PPM	All	All (No Zeros)
SR	-0.04 (0.009) [33]	-0.04 (0.006) [51]	-0.04 (0.005) [54]	-0.08 (0.032) [8]	-0.03 (0.005) [22]	-0.04 (0.005) [84]	-0.04 (0.005) [79]
LR	-0.08 (0.016) [33]	-0.10 (0.012) [51]	-0.09 (0.012) [54]	-0.17 (0.044) [8]	-0.06 (0.013) [22]	-0.09 (0.010) [84]	-0.10 (0.010) [80]
Static	-0.09 (0.019) [31]	-0.06 (0.007) [86]	-0.09 (0.013) [56]	-0.05 (0.007) [48]	-0.03 (0.010) [13]	-0.07 (0.007) [117]	-0.08 (0.008) [97]

The figures are mean, (standard error) and [number of observations]

Table 3 provides some further descriptives that add context in terms of the flow types covered and the prevailing mean levels of AML and GJT. Note that we could source representative AML and GJT data for most but not all of the late time elasticities. The mean GJT figures, which will mainly cover travel time and a wait time element, are clearly larger for inter-urban travel and, across the evidence, cover a large range. The AML figures are of the order of 2½ to 5 min, being larger for longer distance services which are inherently less reliable.

### Summary elasticity statistics

We here present some summary statistics for the assembled elasticity evidence with a focus on those less likely to be impacted by confounding factors, including lesser considered but nonetheless important insights into the length of the long run and the ratio of long run to short run elasticities.<sup>12</sup>

In order to make the PPM elasticities comparable, the same procedure has been followed as in Wardman and Batley (2014). The PPM elasticity is multiplied by the elasticity of PPM to APM to convert it to an (approximate) APM elasticity.<sup>13</sup> The elasticity of PPM to APM is here obtained from regression of the logarithm of PPM on the logarithm of APM using a very large dataset covering 24,574 flows and the years 2009 to 2016. For non-seasons, and the available data of 185,418 observations, the elasticity was -0.15 (t=244), with incremental effects of 0.04 (t=34) for longer distance trips over 20 miles and 0.02 (t=18) for trips to and from London. For season tickets up to 50 miles, and 87,111 observations, the elasticity was -0.13 (t=250). These are very much in line with the -0.11 estimated by Wardman and Batley (2014).

Table 4 provides some summary short run, long run and static late time elasticities. The wrong sign elasticities are converted to zero, of which there are 29 (10%), and the PPM

<sup>12</sup> Whilst we could estimate models along the lines of our main meta-model which explained variations in the length of the long run and the ratio of long run to short run elasticities across the evidence assembled, the number of observations is too few to support the estimation of precise coefficient estimates.

<sup>13</sup> 
$$\eta_{APM} = \frac{\partial V}{\partial APM} \frac{APM}{V} = \frac{\partial V}{\partial PPM} \frac{\partial PPM}{\partial APM} \frac{APM}{V} = \frac{\partial V}{\partial PPM} \frac{\partial PPM}{\partial APM} \frac{APM}{V} \frac{PPM}{PPM}$$

$$= \frac{\partial V}{\partial PPM} \frac{PPM}{V} \frac{\partial PPM}{\partial APM} \frac{APM}{PPM}$$

**Table 5** Ratio of long run to short run elasticities

Constant	2.34	PPM	2.11	Annual	2.09	Seasons	2.75
	(0.24)		(0.28)		(0.35)		(0.53)
	[33]		[22]		[13]		[29]
Proportional	3.12	AML	2.50	Quarterly	1.34	Non-Seasons	2.85
	(0.34)		(0.49)		(0.16)		(0.21)
	[51]	[8]	[6]	[55]			
		APM	3.15	Four Weekly	3.10		
			(0.32)		(0.27)		
			[54]		[65]		

The figures are mean, (standard error) and [number of observations]

**Table 6** Length of long run (in years)

Constant	3.09	PPM	3.46	Annual	3.21	Seasons	2.04
	(0.35)		(0.51)		(0.50)		(0.49)
	[33]		[22]		[13]		[29]
Proportional	0.76	AML	2.31	Quarterly	4.82	Non-Seasons	1.48
	(0.09)		(0.23)		(1.38)		(0.15)
	[51]	[8]	[6]	[55]			
		APM	0.86	Four Weekly	1.08		
			(0.10)		(0.11)		
			[54]		[65]		

The figures are mean, (standard error) and [number of observations]

elasticities are converted into equivalent APM elasticities. Across the various segmentations, the long run elasticities exceed the short run, as is to be expected, although the relationship with the static elasticities varies and may be the result of other unaccounted for influences.

There is little difference between the elasticities derived from the constant elasticity and proportional elasticity functions. In principle, the APM elasticities are expected to be greater than the AML elasticities, because APM additionally contains the cancellations element (DML), but this is not borne out in practice for either the short run or long run, whilst the PPM based elasticities are lower. In these cases, other factors may be at work. The final column removes the zeros and this makes very little difference.

The PDFH long run APM elasticity recommendations reproduced in Table 1 are in the range  $-0.070$  to  $-0.115$ , with the exception of  $-0.245$  for airport flows. Whilst the figures in Table 4 broadly support the PDFH recommendations, we must nonetheless be careful of unaccounted for influences and leave further consideration of the absolute elasticities and their variations to the meta-analysis.

There are 84 instances of ratios of long and short run elasticities from the same study, with a mean of 2.82 and standard error of 0.23. Table 5 presents some variations. The proportional elasticities have a larger ratio as do the APM based elasticities, and this may well be because they are dominated by four weekly data where the short run elasticity can

be expected to be lowest. This is confirmed by the ratio being highest for four weekly data. There is though little difference in the ratio between season and non-season tickets.

Turning to the length of the long run, defined as being 95% of the effect working through, the mean across the 84 observations is 1.68 years with a standard error of 0.19. This average though hides some important variations as is apparent from the results contained in Table 6.

There is a very large difference between the elasticities obtained from constant and proportional functional forms which is repeated for the measure of late time and we suspect this can be attributed to the prevalence of four weekly data in the estimation of proportional elasticity and APM models. Indeed, the elasticities estimated to four weekly data have very much shorter long run effects, although annual and quarterly data form relatively few cases.

As would be expected, the length of the long run is longer for season tickets where constraints surrounding moving house, changing job and the use of annual tickets will imply a longer period of adjustment.

PDFH currently recommends a long run of around three years for non-season tickets and around five years for season tickets, and the higher figures for the constant elasticity and annual data used in older studies may be a contributory factor. Whilst the figures here may be seen to challenge these recommendations, it is though quite alarming that the length of the long run is dependent upon the periodicity of the data used in estimation, particularly given the attractions of using four weekly data in the analysis of late time demand impacts.

## Meta-model

Meta-analysis aims to quantify how a parameter of interest, such as a demand elasticity, varies across the estimates obtained from different studies as a function of the key features of the estimates and studies. Its attractions and limitations have been rehearsed elsewhere (Button 2019; Elvik 2018; Wardman 2012). Our view is that it can provide valuable methodological insights, sometimes not possible by other means, and also estimates of the parameter of interest for situations where evidence does not exist. It also serves to provide a useful summary of existing evidence, against which emerging results can be compared, whilst being more robust to confounding factors and spurious impacts than traditional literature reviews.

Explaining how elasticities vary within and across studies is well suited to regression analysis which is here used. The meta-analysis was conducted using SPSS (IBM 2020). We have the choice of estimating an additive or a multiplicative model to explain the late time elasticities ( $\eta$ ) here expressed in absolute form. The Wardman and Batley (2014) meta-analysis of late time elasticities used the additive form whereas various other meta-analyses of valuations of time, price elasticities, time elasticities and cross-elasticities have used the multiplicative form. The extent of testing the two forms has been limited.

The additive model would be specified as:

$$\eta = \alpha + \beta X + \dots \quad (5)$$

Here the  $\beta$  parameters denote an additive effect on  $\eta$  from amongst the set of explanatory variables ( $X$ ) listed in “Data assembly” section that we collected evidence for. The multiplicative model would be specified as:

**Table 7** Meta-model

	Coeff (t)	Effect (%)
Constant	-2.918 (6.1)	
Base = short run four weekly		
Short run annual (n = 13)	0.711 (3.3)	+ 104
Static annual (n = 65)	0.682 (4.8)	+ 98
Long run (n = 84)	1.144 (8.3)	+ 214
Base South East to/from London		
Non-London Urban (n = 33)	-1.793 (8.7)	- 83
London TCA (n = 19)	-0.581 (3.6)	- 44
Airports (n = 10)	0.290 (1.7)	+ 34
Base = Urban		
Inter-Urban (n = 161)	-1.490 (9.4)	- 77
Urban and Inter-Urban (n = 67)	-1.015 (6.3)	- 64
Base = Non seasons full fare		
Seasons (n = 85)	-0.328 (2.6)	- 28
Base = APM		
PPM (n = 57)	-0.817 (4.8)	- 56
Base = study purpose not late time		
Study purpose late time (n = 123)	-0.889 (6.9)	- 59
Base = Not significant at 30%		
Significance $\geq 10\%$ (n = 218)	1.437 (3.2)	+ 321
Base = not study A		
Study A (n = 13)	0.567 (5.0)	+ 76
Adjusted R <sup>2</sup>	0.539	
Observations	285	

$$\eta = e^{\alpha + \beta X + \dots} \quad (6)$$

The  $\beta$  parameters here denote a multiplicative effect on  $\eta$ . The usual means of estimating Eq. 6 would involve a logarithmic transformation to yield:

$$\ln(\eta) = \alpha + \beta X + \dots \quad (7)$$

However, a problem here is that  $\eta$  can be zero and therefore cannot be included in Eq. 7. Equation 6 is therefore estimated using non-linear least squares whereupon the zero elasticities can be included. This also has the attraction that the sum of squares measures of Eqs. 5 and 6 are directly comparable.

For the same independent variables, directly estimated Eq. 6 achieved somewhat lower residual sum of squares whilst also providing larger t ratios for almost all coefficient estimates. We have therefore proceeded with this multiplicative model.

Table 7 reports the meta-model resulting from the examination of the variables set out in “Data assembly” section and retaining those coefficients which are significant at the usual 5% level or were deemed to merit retention. The base category is specified and the number of observations relating to each estimated parameter is given. The proportionate effect of a parameter estimate relative to the base is also reported.

Table 7 contains an encouraging number of significant effects, and indeed contrasts with Wardman and Batley (2014) where the reported models contained only three. Many of the coefficient estimates are estimated very precisely with significance at the 0.1% level or better. The goodness of fit is respectable given the diverse nature of the studies and the challenges involved in estimating late time elasticities. As for correlations between the various coefficient estimates, only those between Inter-Urban and Non-London Urban (0.69), Urban and Inter-Urban and Inter-Urban (0.77) and Static Annual and Long Run (0.68) exceeded 0.6.

### Dynamic effects

Given that the long run, static and particularly the short run elasticities will depend upon the temporal resolution of the data used in estimation, the following interaction variables were specified:

- Short Run Four Weekly (n=65)
- Short Run Quarterly (n=6)
- Short Run Annual (n=13)
- Long Run Four Weekly (n=65)
- Long Run Quarterly (n=6)
- Long Run Annual (n=13)
- Static Four Weekly (n=14)
- Static Quarterly (n=38)
- Static Annual (n=65)

Admittedly, this is spreading the data a little thinly and we would not expect to detect significant effects for all segments even if they were materially different.

The short run four weekly elasticity is the base category. The short run quarterly elasticity term was not significant, presumably as a result of the few cases. Short run annual data was found to return late time elasticities that were just over twice the four weekly equivalents. Whilst a larger ratio might be expected, given that an annual short run covers thirteen four weekly periods, the use of annual data might be expected to dampen the effect obtained given the averaging involved.

The only significant effect obtained for the static terms was for annual data, although this has the most cases. It indicates late time elasticities almost twice the short run four weekly elasticities.

As for long run elasticities, the four weekly, quarterly and annual terms yielded broadly similar effects. The three categories were therefore combined into a single term which implies late time elasticities 3.14 times larger than short run four weekly elasticities.

### Flow type and distance

PDFH elasticity recommendations distinguish by flow type and distance category for reasons of the historical organisation of the railways, the evidential basis and the expectations of elasticity variation. The segments are set out in Table 1 and were followed here.

The base category was specified to be South East to and from London TCA, an inter-urban market where rail is in a strong competitive position. Relative to this, elasticities within the London TCA are 44% lower and within Non-London urban areas 83% lower.



The TCA results must be treated with some caution, for reasons already discussed, but it is plausible that urban travellers outside of the South East are less concerned about late time and could possibly be more captive to rail. Distance effects were apparent, with inter-urban trips over 20 miles indicating a large reduction in elasticity of 77%, and this is taken to represent late time being more expected and hence tolerated on longer distance journeys. Elasticities which did not distinguish between urban and inter-urban trips were found to be 64% lower. The effect for airport travel is only significant at the 10% level but is retained since we would expect sensitivity to be larger in this market. It is 34% larger than the base although such flows will also tend to be longer distance.

### **Ticket type**

Late time elasticities might be expected to vary by ticket type given season tickets cover commuting trips, full fare tickets are more often used by business travellers, and off-peak and discounted tickets are attractive to leisure travellers. However, the only significant effect was that season ticket late time elasticities are 28% lower. This can be attributed to commuters being more captive to rail. Whilst this might be considered a relatively small effect, commuters will have much greater awareness of performance than other less frequent users.

### **Measure of late time**

We have three measures of late time as previously described. APM was specified as the base category. There was no significant difference for AML elasticities, but this is not surprising since AML forms the largest part of APM, and AML and APM tend to be highly correlated. Elasticities based on PPM were though 56% lower. This could be a function of the approximations involved in our conversion of PPM elasticities to equivalent APM elasticities, but other contributory factors are that the PPM elasticities are less granular in terms of more often being annual and specified at a coarser spatial detail. It is though important to have isolated this effect.

### **Study and data quality**

A criticism of meta-analysis is that all evidence is treated as equally reliable. There are several means by which this concern can be addressed:

- Examine whether the late time elasticity varies with the source of the evidence and when the study was conducted.
- Explore whether and to what extent the estimated effects in the meta-model, and their associated standard errors, depend upon the precision with which each observation was originally estimated.
- Allow for estimation method and data used and indeed the purpose of the study, on the grounds that 'better quality' evidence will be obtained from more advanced methods and better data, and these effects may be amplified where the explicit purpose of the study was to estimate late time elasticities.
- Include study-specific effects as a means of identifying 'poor quality' studies.
- Remove 'outliers' from the estimated meta-model.

We discuss our examination of each of these issues in turn.

**Source of evidence** Whilst evidence published in peer-reviewed academic journals might be expected to be of higher quality, this is represented by only one of our 17 studies with the remainder reports to the railway industry or government bodies. We would though point out that our previous meta-analyses of values of time and of demand elasticities detected very limited effects from the source of the evidence.

**Study period** There might be a view that more recent studies provide more reliable evidence, not least because performance data and estimation methods have improved over time. There were no significant effects when specifying variables representing the time period of reporting or the time period of the data used in estimation. However, all but one study reported in the past 15 years with almost three-quarters since 2010. The estimation dataset is similarly narrow with older data having a large proportion of PPM elasticities and recent studies covered by the study purpose which both have their own terms in the meta-model.

**Precision of assembled evidence** Evidence was collected on two variables which are of use here: the t ratio of the estimated late time parameter and the number of observations used in estimation.

The number of observations (NOBS) can be taken as a proxy for the reliability of elasticity estimates and was used in a weighted estimation procedure using a weight (W) function of:

$$W = \frac{1}{NOBS^4} \quad (8)$$

The estimation procedure, available within SPSS, searches for the best fit across a set of  $\lambda$ , here ranging from  $-2$  to  $+2$  in units of  $0.05$ . The best fit was achieved at a  $\lambda$  of  $0.2$  and it made virtually no difference to the coefficient estimates and t ratios of the meta-model. Hence the weighting is not used.

As for the t ratio evidence, in 45 (16%) of cases the study only reported whether the late time elasticity estimate was significant at the 10%, 5% or 1% level.<sup>14</sup> We have therefore explored the impact of the significance level categorisations in Table 2 on the late time elasticity. The pattern of results was not monotonic but it indicated that late time elasticities were, unsurprisingly, larger when the significance level was 10% or better. Three-quarters of observations meet this criterion and they yield elasticities that are on average 4.2 times larger. The issue then is whether the non-significant estimates reflect late time elasticities that are very low or there are factors at play which mean the estimation has been unsuccessful. We return to this point in “[Implied late time \(APM\) elasticities](#)” section.

**Estimation method, data used and study purpose** The estimation method, including the two-way fixed effects used in recent studies, did not have an effect on the late time elasticities. Whilst it can be hypothesised that the late time data used might have an impact, given that more disaggregate (i.e. spatially/temporally) late time measures would be of higher quality, no significant variations were apparent on this account. However, there will be cor-

<sup>14</sup> Even though significance at 1% indicates a precise estimate, there is a difference between a t ratio of, say, 3 and 13, both of which indicate significance at the 1% level.

relation between the use of flow-specific data and the purpose of the study being specifically to estimate late time elasticities since these paid more attention to the specification of late time, to the collection of bespoke late time data, to the judicious selection of flows to be analysed and to the specification and analysis of late time. A significant effect was obtained for a variable denoting that the purpose of the study was late time elasticity estimation, of which there are three such studies, and on average the elasticities are 59% lower. This we would contend is a quality-related effect.

**Study-specific effects** Previous meta-analyses have used study-specific effects in order to detect poor quality studies which systematically recover parameters that are too low or high. There are though a number of drawbacks of this approach: when using the meta-model to forecast we need to decide whether the study-specific effects are genuine or not, which can make a big difference; the study-specific effects may weaken effects that should be attributed to the explanatory variables, particularly when a study yields relatively few observations; and not all evidence from a specific study is necessarily of poor quality.

The approach adopted here is to examine the residuals to determine whether some studies have provided elasticities which cannot be explained and then to consider whether there are reasons for this. Observations with standardised residuals outside the range  $\pm 1.96$ , which will represent 5% of the total, were inspected. This is in line with the use of 5% significance levels more generally. One study had three such outliers and six had two outliers but these were small proportions of the number of observations in each study. The former study (Study A) had a consistent pattern of positive residuals, although for no clear reason, and therefore a study-specific effect was entered for it which indicates that its late time elasticities are on average 76% larger than for other studies, all else equal. The inclusion of this study-specific effect had hardly any impact on the other coefficient estimates.

**Remove outliers** When the 5% of observations with standardised residuals outside the range  $\pm 1.96$  were removed, which might be considered to represent the poorest quality and hardest to explain evidence, the goodness of fit increased somewhat to 0.647 but there was relatively little impact on the coefficient estimates whose absolute deviation averaged 21%.

### Summary of quality-related issues

The results here seem to indicate that quality-related issues are not impacting greatly on the results, although it could reasonably be hypothesised that variations in study quality have a random impact on the outturn elasticity estimates. Unsurprisingly, the main effect is attributable to whether the late time elasticity estimate was significant at the 10% level or better.

### Implied late time (APM) elasticities

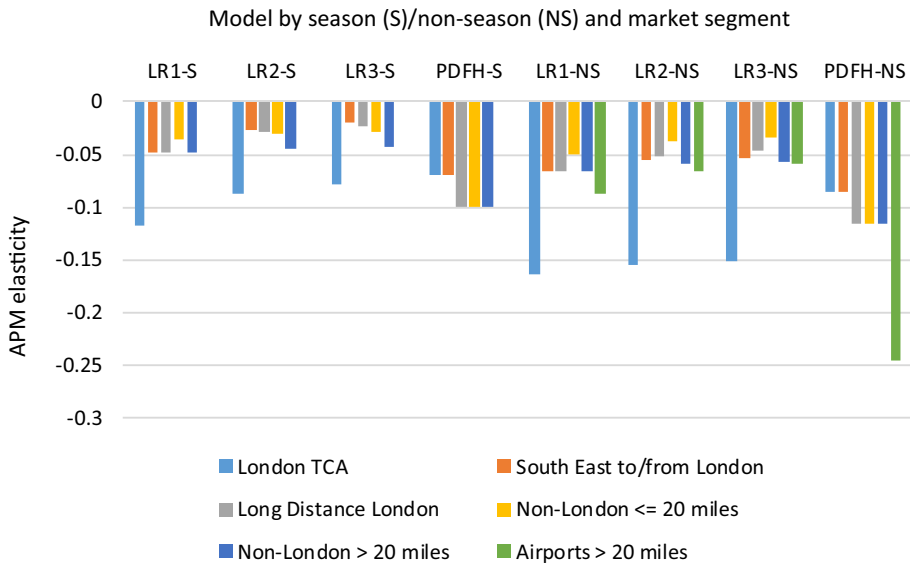
Table 8 provides implied long run APM late time elasticities for the various PDFH segments along with PDFH recommendations and the results of the recent Steer (2019) study as given in Table 1. Figure 1 presents the same results in graphical form.

The meta-model contains a term for late time elasticity estimates that were significant at 10% or better. These are obviously somewhat larger than those elasticities that were not significant at this level. However, to ignore the latter means that late time elasticities that are not significant at the 10% level or better are implicitly assumed to be the same as those that are. This would seem to be highly unlikely.

**Table 8** Implied long run late time (APM) elasticities

Scenario	% INSIG	LR1	LR2	LR3	PDFH	Steer (2019)
London TCA						
Season	0.33	-0.118	-0.088	-0.079	-0.070	n.s
Non-season	0.08	-0.164	-0.154	-0.151	-0.085	n.s
South East to/from London						
Season	0.58	-0.048	-0.027	-0.020	-0.070	n.s
Non-season	0.20	-0.066	-0.056	-0.053	-0.085	-0.033
Long Distance London						
Season	0.50	-0.048	-0.029	-0.024	-0.100	n.s
Non-season	0.31	-0.066	-0.051	-0.046	-0.115	-0.035
Non-London ≤ 20 miles						
Season	0.16	-0.035	-0.031	-0.029	-0.100	n.s
Non-season	0.33	-0.049	-0.037	-0.033	-0.115	-0.088 and -0.096
Non-London > 20 miles						
Season	0.11	-0.048	-0.044	-0.043	-0.100	-0.041
Non-season	0.14	-0.066	-0.059	-0.057	-0.115	-0.029 and -0.050
Airports > 20 miles						
Non-season	0.33	-0.088	-0.066	-0.059	-0.245	n.s

Implied elasticities are for where the study purpose was specifically to estimate late time elasticities. All figures are APM



**Fig. 1** Implied long run late time (APM) elasticities

As discussed in “[The problems of estimation and policy implications](#)” section, insignificant estimates might simply reflect a very low elasticity in practice or might stem from difficulties in estimation, and unfortunately we have no clear means of distinguishing between the two. We therefore provide three sets of implied elasticities.

- LR1 which ignores the insignificant late time elasticities. This provides an upper bound.
- LR2 which uses the meta model’s implied late time elasticities for insignificant observations. These are not zero but are only around a quarter of those implied where elasticities are significant at the 10% level.
- LR3 which takes the insignificant estimates to be zero. This provides a lower bound.

Table 8 therefore provides the proportion of late time elasticities in our dataset that are not significant (% INSIG) at the 10% level for each flow and ticket type combination. Where rail is in its strongest competitive position, for season tickets in the South East and long distance London trips, which can be expected to and indeed does drive low elasticities, the proportion of insignificant elasticities is highest. The comparatively small number of airport flows could well have contributed to the relatively large number of estimates that were insignificant in this market. The small proportion of Non-London long distance flows with insignificant estimates presumably stems from generally very large estimation datasets.

Apart from London TCA elasticities, which could be influenced by the demand data problems discussed in “[Key recent insights](#)” section, each of the three elasticities implied by the meta-model are lower than PDFH recommendations, with LR1 being on average a half of the PDFH figure falling to around a third for LR3.

In three out of four cases, our implied elasticities are also larger than the significant estimates obtained by Steer (2019), and including the latter study’s insignificant estimates our implied elasticities would be larger in 10 out of 11 cases, generally lying between the PDFH and Steer (2019) figures.

Given that LR1 provides an upper bound and LR3 provides a lower bound, and LR2 is not only between the two but is based on low but non-zero late time elasticities implied by the meta-analysis for those instances where the estimated elasticities were not significant, our pragmatic preference is for the LR2 figures. With the exception of London TCA, we find these implied late time elasticities and their pattern of variation to be credible.

The implied elasticities challenge the prevailing PDFH recommendations based on Wardman and Batley (2014), although not to the extent of Steer (2019). A key difference is the presence here of an effect relating to the purpose of the study being to estimate late time elasticities which reduces the implied elasticities by 59%. The retention and use of this term in calculating implied late time elasticities seems justified for reasons already discussed. Nonetheless, given that 43% of the observations in the model are from studies whose specific purpose was to estimate late time elasticities, ignoring the effect would increase the implied elasticities in Table 8 by a factor of 1.82 and this would still leave the implied LR2 elasticities generally less than the PDFH recommendations and for the non-London TCA flows on average 27% less.

## New econometric insights

We here present some new econometric insights, driven in part by the outcomes of recent studies and focussed upon variations in late time elasticities. The data available to us has attractions in that it is very large, covers many different types of flow and its previous analysis (Leigh Fisher et al. 2016) forms the basis of current PDFH recommendations regarding the effects of external factors. It is annual station-to-station demand covering the years between 1995/96 and 2013/14 split by season and non-season tickets. It was provided by the Department for Transport who assembled it from revenue, demand and GJT data supplied by the train companies, reliability data sourced from the infrastructure provider (Network Rail) and official statistics.

The measure of lateness available is AML for origin station to destination station flows.<sup>15</sup> The railways in Great Britain have in place a detailed recording system of train performance based on monitoring points along routes. This is the industry standard for the quantification of AML. The AML data used here is derived from the PEARS database which records lateness for service codes, which are groupings of services based on train service patterns, with weighted averages taken where a flow is served by more than one service code. Since the AML data was only available from 2002/3, and for some flows only for the final six years, it did not feature in the models reported in Leigh Fisher et al. (2016). Moreover, that study identified some shortcomings of the AML data supplied, largely related to service code changes and recording errors.

We have re-examined the AML data and identified the structural breaks, often linked to new services, and removed the offending data – such that we now believe that the data is fit for purpose. In some cases, we have removed flows entirely, as in what were then the Thameslink, Southern and Great Northern franchises because they had a break in 2007 and it was not clear whether the pre- or post-2007 figures were reliable. In other cases, we have only removed some years which were clearly inaccurate, such as outer suburban services on the West Coast Main Line to London where up to 2008 the AML figures were implausibly low but thereafter exhibited credible values and patterns.

Our examination of the data identified that large structural breaks in AML most commonly occurred on flows within the South East area and when these observations were removed the AML coefficient for both season and non-season models were no longer significant. Nor were we able to estimate significant late time parameters for long distance London flows. Models are reported for the following PDFH flow and ticket type categories:

- Non-London non-season tickets for shorter distances up to 20 miles (NSS)
- Non-London non-season tickets for longer distances over 20 miles (NSL)
- Non-London season tickets up to 20 miles (SS)
- Non-London season tickets for 21–75 miles (SL)<sup>16</sup>

<sup>15</sup> Whilst PDFH uses APM for forecasting, and we only had AML figures available, generally AML accounts for around three-quarters of APM. Moreover, we avoid the assumptions necessary in creating DML. Batley et al. (2011) found little difference between elasticities based on AML and APM as did the meta-analysis in “Meta-model” section. We therefore conclude that AML is a reasonable representation of train performance.

<sup>16</sup> The amount of commuting tapers off very sharply on Non-London flows over 75 miles. Our subsequent models were not particularly sensitive to using 50 miles, which was used in Leigh Fisher et al. (2016), or 75 miles as the upper limit.

Flows involving interchange were removed, in line with the recent OXERA and Winder Phillips (2017) and Steer (2019) studies, because AML is not directly recorded for such flows<sup>17</sup> whilst the removal or introduction of interchange could itself lead to structural breaks in the AML data.

Table 10 below illustrates the levels of AML for these four flow types. The mean levels of GJT were 38, 165, 39 and 69 min with average distances of 10, 87, 11 and 34 miles respectively. Demand levels are substantial in all cases, with mean station-to-station trips per annum of approximately 18, 10, 15 and 12 thousand respectively and 25th percentiles of around 9, 3, 8, and 7 thousand.

The dependent variable is in all models the logarithm of the annual volume (V) of demand between stations. The specification of the independent variables varies across the different fixed effects models. The models were estimated using SAS (SAS Institute 2020).

All models contain: fare, specified as revenue per trip; GJT; car cost, including efficiency improvements; and car journey time, based upon speed measures derived from historical National Travel Survey data. These also enter in logarithmic form and hence their parameters are interpreted as elasticities. Time trends were specified which indicate the proportionate annual change in rail demand due to factors not represented in the models.

Season ticket demand is made a function of employment at the destination, specified at district level and entered in constant elasticity form. Given season ticket demand is driven by employment, population is not entered. Non-season ticket demand is made a function of origin population, specified at district level, and GVA per capita, specified at sub-regional (NUTS3) level. These also take the constant elasticity form.

The cross-elasticities to car cost and car time, the population elasticity for non-seasons and the employment elasticity for seasons have been constrained to best evidence, in line with previous research (Leigh Fisher et al. 2018; Wardman 2006; Wardman et al. 2019). This is because these tend to be correlated amongst themselves and with other variables whereupon the freely estimated parameters can be implausible with implications for other parameter estimates. Note also that the employment and population terms allow for differential propensities to make rail trips across different age groups, car ownership levels, occupation types and employment sectors as set out in Leigh Fisher et al. (2016).

Of particular interest here is the specification of the AML variable and a range of formulations with different elasticity ( $\eta$ ) properties have been explored. AML enters our models in constant elasticity form as:

$$\ln V = \alpha \ln AML \quad \eta = \alpha \tag{9}$$

For brevity of presentation, terms other than AML are not here shown. The AML elasticity is directly proportional to the level of AML if the demand model is specified as:

$$\ln V = \beta AML \quad \eta = \beta AML \tag{10}$$

A more flexible function is to enter AML as:

$$\ln V = \gamma AML^\lambda \quad \eta = \gamma \lambda AML^\lambda \tag{11}$$

As  $\lambda$  tends to 0 (1) then the AML elasticity tends to the constant (proportional) form.

We have here gone beyond previous work in exploring additional functional forms. The PDFH approach to forecasting the impact of AML variation on rail demand prior to 2018

<sup>17</sup> In other words, AML data records the movement of trains not people.

was the indirect method set out in “Forecasting changes in late time in the rail industry in Great Britain” section. This was a pragmatic approach to extending the GJT forecasting approach to include late time in an era when late time data for inclusion in econometric models was not available in the volume and detail that it is today. We can now explore an explicitly extended GJT term (EGJT) as follows<sup>18</sup>:

$$\ln V = \theta \ln EGJT \quad EGJT = GJT + \varpi AML \quad \eta = \theta \frac{\varpi AML}{EGJT} \tag{12}$$

where  $\varpi$  represents the lateness multiplier. In principle, the inclusion of AML to create the larger EGJT term should result in an elasticity to EGJT that is larger than to GJT. An innovation here is to directly estimate  $\varpi$  which provides a grounding in the Revealed Preference (RP) of actual behaviour and avoids concerns that have been raised about the large valuations of  $\varpi$  that can often emanate from SP studies.

A downside of Eq. 12 is that it forces the AML elasticity to be directly related to the proportion that AML forms of EGJT. This is a strong assumption and one that should be subject to empirical testing. A more flexible relationship would be obtained by specifying:

$$\ln V = \theta EGJT^\tau \quad \eta = \theta \tau \frac{\varpi AML}{EGJT^{1-\tau}} \tag{13}$$

As  $\tau$  approaches 1 the dependence of the AML elasticity on the level of GJT diminishes. The form that AML enters could also be varied to dampen the relationship between  $\eta$  and AML. However, a limitation of Eq. 13 is that the dependence of the elasticities of all of the other variables within GJT will have the same modified relationship in terms of  $1 - \tau$ . We therefore do not pursue this function. Instead, Eq. 11 has been extended to explicitly formulate the AML elasticity alone as a function of the proportion it forms of GJT, as follows:

$$\ln V = \mu \frac{AML^\lambda}{RGJT^\gamma} \quad \eta = \mu \lambda \frac{AML^\lambda}{RGJT^\gamma} \tag{14}$$

This function uses the reference level of GJT (RGJT), which is the level of GJT for each flow in 2014. Given that RGJT does not vary across the different observations within a flow, it only enters the elasticity functions as a modifier. If the GJT denominator term were allowed to vary, then this would impact on the GJT elasticity itself and complicate matters in terms of the interpretation of the results.

It can be anticipated that the estimation of Eq. 14 will be challenging, given the presence of three parameters for the AML term. Our final model is therefore a simple extension of Eq. 10:

$$\ln V = \mu AML + \kappa \frac{AML}{RGJT} \quad \eta = \mu AML + \kappa \frac{AML}{RGJT} \tag{15}$$

Table 9 reports models based on Eqs. 9, 10 and 11. Reassuringly, the fare, GJT, GVA and time trend estimates vary little according to the formulation of AML. The fare and GJT elasticities are very precisely estimated and plausible, which is encouraging. An effect due

<sup>18</sup> In addition, it is common practice in transport modelling and forecasting to use composite terms, and indeed GJT in its current form is one such example, and so examination of this enhanced terms is warranted. Whilst the  $\gamma$  and  $\varphi$  used in the computation of GJT could also be directly estimated, as for example in Wheat and Wardman (2017), we have here for consistency purposes used the PDFH recommended values.



**Table 9** Estimation results for Eqs. 9–11

Equation	NSS: Non-London Non-Seasons ≤ 20 miles			NSL: Non-London Non-Seasons > 20 miles			SS: Non-London Seasons ≤ 20 miles			SL: Non-London Seasons 21–75 miles		
	9	10	11	9	10	11	9	10	11	9	10	11
In Fare	-1.05 (59.7)	-1.02 (58.6)	-1.02 (58.0)	-0.64 (48.4)	-0.64 (48.1)	-0.64 (47.9)	-0.84 (32.3)	-0.84 (32.3)	-0.84 (32.3)	-0.38 (7.2)	-0.38 (7.3)	-0.37 (7.1)
In GJT	-0.76 (31.5)	-0.75 (31.3)	-0.75 (31.3)	-0.73 (36.3)	-0.73 (36.3)	-0.73 (36.5)	-0.62 (8.6)	-0.62 (8.6)	-0.62 (8.7)	-0.97 (7.8)	-0.98 (8.0)	-0.94 (7.8)
In GVA	0.48 (20.9)	0.41 (17.6)	0.39 (16.4)	0.63 (19.4)	0.63 (19.5)	0.63 (19.7)	0.030 (19.7)	0.030 (19.7)	0.030 (19.7)	0.022 (12.9)	0.022 (12.9)	0.020 (11.2)
Trend	0.019 (37.0)	0.017 (34.5)	0.018 (34.9)	0.030 (57.6)	0.030 (57.0)	0.030 (57.1)	0.022 (16.4)	0.021 (15.6)	0.021 (16.2)	0.022 (12.9)	0.021 (11.7)	0.020 (11.2)
WC Disrupt				-0.11 (10.4)	-0.12 (10.8)	-0.12 (10.7)						
In AML	-0.066 (17.3)			-0.010 (2.8)			-0.001 (0.1)			-0.008 (0.6)		
AML		-0.033 (20.9)	-0.0082 (3.7)		-0.0052 (4.3)	-0.0003 (1.1)		-0.012 (2.3)	-0.00006 (0.5)		-0.013 (2.9)	-0.0005 (0.8)
λ		1.73 (12.4)	1.0		2.40 (5.7)	1.0		4.07 (3.8)			2.63 (3.8)	
In Population	1.0	1.0	1.0	1.0	1.0	1.0						
In Employment							1.0	1.0	1.0	1.0	1.0	1.0
In Car Cost	0.40	0.40	0.40	0.26	0.26	0.26	0.40	0.40	0.40	0.40	0.40	0.40
In Car Time	0.20	0.20	0.20	0.30	0.30	0.30	0.20	0.20	0.20	0.20	0.20	0.20
RSS	383.03	380.60	380.21	213.33	213.12	212.86	328.08	327.80	326.89	172.21	171.78	171.12
Adjusted R <sup>2</sup>	0.985	0.986	0.987	0.990	0.990	0.991	0.939	0.940	0.940	0.938	0.938	0.939
Flows	2960			1994			957			454		
Observations	23,988			13,085			6990			3547		

to the disruptions associated with the major upgrade to the West Coast Main Line (WC Disrupt) of the order of 12% per annum was discerned but other effects due to new rolling stock, revenue protection, on-track competition and ticket type switching were not significant. The GVA elasticities for non-season tickets are relatively low, although the time trend terms may have discerned some of the demand effects of GVA.

The time trends indicate a 2–3% annual growth in demand and have very large  $t$  ratios. Wardman and Lyons (2016) pointed out that the ability to do more and better worthwhile activities while travelling due to the digital revolution might be expected to impact on rail demand and Wardman et al. (2020) demonstrated such an effect. In addition, during the period under investigation there were structural changes in the labour market towards white collar jobs and consolidation of employment in regional centres well served by rail. These cannot be explicitly accounted for in the modelling but can be expected to have created strong trends in rail demand.

The data supports the estimation of robust models with credible parameters which provides a sound basis for the investigation of the demand impacts of AML. It can be seen that the proportional elasticity model achieves the better fit and provides more precisely estimated late time parameters. Indeed, restricting analysis to the constant elasticity form would lead to the removal of the lateness variable from two models as insignificant.

The variation in elasticities across the different market segments seems credible. For short distance trips, the effect is smaller for commuters and this is attributed to rail's stronger competitive position in the peak and the mandatory nature of commuting. For the longer distance non-season trips, it is not unreasonable to speculate that greater unreliability is expected and therefore tolerated.

The issue of removing the flows and observations with 'suspect' AML data might be deemed controversial. When these observations are restored, the AML parameters and  $t$  ratios of Eq. 10 are respectively  $-0.029$  (23.0),  $-0.007$  (5.0),  $-0.042$  (13.3) and  $-0.016$  (4.1), which apart from the SS model are not greatly different. However, the presence of the time trend does have an appreciable effect; its removal leads to coefficient estimates of  $-0.059$  (41.3),  $-0.013$  (9.5),  $-0.043$  (8.5) and  $-0.028$  (7.0) which are between 1.8 and 3.6 times larger. It is possible that the time trends have discerned some of the late time effect. Nonetheless Steer (2019), whose models also contained inter-temporal effects, reported long run figures of around  $-0.045$  for NSS and between  $-0.011$  and  $-0.020$  for NSL, which are both larger than here obtained, and a figure for SL which is a little larger at  $-0.016$  although they could not recover a significant estimate for SS.

Equation 10 was estimated with AML lagged one year. The lagged term was not significant in the NSS, NSL and SL models whilst it was significant but wrong sign for the SS model. It could be that most of the effect works through within a year, although we note the comment in Wardman et al. (2019) that recent rail demand models have struggled to estimate credible lagged effects.

The final model in Table 9 is the more flexible Eq. 11. In all cases the  $\lambda$  parameter is statistically significant and exceeds 1 by a considerable margin, albeit with the  $\beta$  parameter significant only for the NSS model. This means that the marginal impact of a change in AML becomes greater as AML increases. These  $\lambda$  estimates here contrast with those reported by Steer (2019) which lay between 1 and 2 and with the squared terms in the quadratic functions estimated by OXERA and Winder Phillips (2017) which were small and insignificant. Whilst we have to accept that the  $\beta$  parameter is only significant for the NSS model, the precise estimation of the  $\lambda$  parameters which are all significantly different from one should not be ignored. We initially had sympathy with the view expressed by Steer (2019) that, "Intuitively, the relationship between change in AML and revenue

might be expected to sit somewhere on the range between the two extremes of constant and semi-elasticity". However, given both our findings and those of Steer (2019) point to  $\lambda s$  in excess of 1, there could be a very important and overlooked factor at work here. It may be that small amounts of late time are not perceived by some rail travellers and even when perceived might not cause much inconvenience. However, explicit analysis of such effects is beyond the scope of this study.

Table 10 provides the AML elasticities implied by Eqs. 10 and 11 for various percentiles of AML in the estimation data. Large proportionate variations in AML elasticities are implied.

Table 11 reports the estimations of Eqs. 12, 14 and 15 which are concerned with how late time fits within the GB railway industry's GJT formulation.

The elasticity to EGJT of Eq. 12 is larger than the elasticity estimated to GJT of Eq. 9 as expected, although generally not by a large magnitude which is consistent with the low impacts of AML on demand here estimated. Significant estimates of the late time multiplier ( $\varpi$ ) were obtained for all but the SS model. What is noticeable is that the 'demand consistent'  $\varpi$  are somewhat lower than existing evidence. Even when the time trends were removed, the estimated  $\varpi$  remained low, and were respectively (with t ratios) 2.57 (23.7), 1.37 (10.5), 1.75 (6.0) and 1.16 (5.0). Compare the estimates of  $\varpi$  in Table 11 with those of the Wardman and Batley (2014) meta-analysis of British SP evidence which ranged between 2.3 and 5.1, depending upon purpose and distance. From its inception in 1986, PDFH used a  $\varpi$  of 2.5 increasing to 3.0 in version 4.0 in 2002 and with the most recent figures in version 5.1 in 2013 being 3.0 and 2.3 for short distance seasons and non-seasons respectively and 3.9 and 3.4 for long distance seasons and non-seasons respectively.<sup>19</sup> This implies that using available  $\varpi$  evidence, at least that drawn from SP studies, is not appropriate. This was the conclusion of the Wardman and Batley (2014) review although obtained differently by comparing directly estimated AML elasticities with those implied by the indirect approach.

Table 12 provides the AML elasticities implied by Eq. 12. Compared to Table 10 and the elasticities implied by the proportional elasticity models of Eqs. 10 and 11, the results here imply less elasticity variation with the exception of the NSS flows where it is broadly similar. The proportion that AML forms of EGJT would not seem to be a key driver of the AML elasticity. However, when we inspect whether EGJT or separate GJT and AML terms provide a better fit, the comparison of the RSSs of Eqs. 9 and 12 for the same number of estimated parameters and constant elasticity function is mixed. Equation 9 provides the better fit for NSL and SS with the reverse for NSS and SL. Given these mixed results, and indeed the pattern of implied elasticities here compared to the Eqs. 10 and 11 which are based solely on AML, further analysis of the dependency of the AML elasticity on the level of AML and GJT is warranted in terms of Eqs. 14 and 15.

As for Eq. 14, it can be seen that the  $\lambda$  terms are essentially demonstrating the same effect as in Eq. 11. As for the  $\gamma$  parameter that drives the variation with GJT, it indicates that the relationship imposed by the enhanced GJT approach, and by implication for the previous PDFH method, is justified for the NSL model, and a weaker effect is supported for the NSS model but for season tickets there is no relationship. However, we again find that the main parameter ( $\mu$ ) is only significant for the NSS model, which is not surprising given the challenges of estimating three parameters to AML.

<sup>19</sup> These continue to underpin Schedules 4 and 8 for non-LSE flows.

**Table 10** Implied AML elasticities

Percentiles	NSS		NSL				SS				SL					
			Equation 11		Equation 10		Equation 11		Equation 10		Equation 11		Equation 10		Equation 11	
	AML	Equation 10	Equation 10	Equation 11	AML	Equation 10	Equation 11	AML	Equation 10	Equation 11	AML	Equation 10	Equation 11	AML	Equation 10	Equation 11
10%	0.93	-0.03	-0.01	1.24	-0.01	-0.01	-0.00	1.07	-0.01	-0.00	1.37	-0.02	-0.00	1.81	-0.02	-0.00
25%	1.33	-0.04	-0.02	1.81	-0.01	-0.01	-0.00	1.34	-0.02	-0.00	1.81	-0.02	-0.00	2.46	-0.03	-0.01
50%	1.69	-0.06	-0.04	2.59	-0.01	-0.01	-0.01	1.71	-0.02	-0.00	2.46	-0.03	-0.00	3.45	-0.04	-0.01
75%	2.20	-0.07	-0.06	3.97	-0.02	-0.02	-0.02	2.30	-0.03	-0.01	3.45	-0.04	-0.01	4.43	-0.06	-0.03
90%	3.03	-0.10	-0.10	5.32	-0.03	-0.04	-0.04	3.05	-0.04	-0.02	4.43	-0.06	-0.02			

**Table 11** Estimation results for Eqs. 12, 14 and 15

Equation	NSS: Non-London Non-Seasons ≤ 20 miles			NSL: Non-London Non-Seasons > 20 miles			SS: Non-London Seasons ≤ 20 miles			SL: Non-London Seasons 21–75 miles		
	12	14	15	12	14	15	12	14	15	12	14	15
Log Fare	-1.03 (58.5)	-1.02 (57.9)	-1.03 (58.4)	-0.63 (48.5)	-0.64 (48.1)	-0.65 (48.1)	-0.84 (32.3)	-0.84 (32.3)	-0.85 (32.4)	-0.38 (7.2)	-0.37 (7.1)	-0.37 (7.1)
Log GJT		-0.75 (31.2)	-0.75 (31.2)		-0.73 (36.5)	-0.74 (36.4)		-0.62 (8.6)	-0.62 (8.7)		-0.95 (7.7)	-0.96 (7.9)
Log EGJT	-0.82 (34.0)			-0.74 (35.9)			-0.64 (8.6)			-1.00 (8.0)		
Log GVA	0.42 (18.0)	0.38 (15.9)	0.40 (17.2)	0.65 (20.3)	0.63 (19.5)	0.63 (19.5)						
Trend	0.018 (36.1)	0.018 (35.2)	0.017 (34.4)	0.031 (60.0)	0.030 (57.0)	0.030 (56.6)	0.022 (16.0)	0.021 (15.8)	0.021 (15.6)	0.021 (11.9)	0.020 (11.2)	0.021 (11.6)
WC Disrupt				-0.12 (10.9)	-0.12 (10.8)	-0.12 (10.8)						
$\omega$	1.55 (15.5)			0.79 (4.4)			0.43 (1.3)			0.54 (2.0)		
$\mu$		-0.035 (2.5)	-0.020 (6.3)		-0.026 (0.9)	0.002 (0.9)		-0.00002 (0.5)	-0.047 (2.9)		-0.00007 (0.4)	-0.048 (3.3)
$\lambda$		1.76 (12.1)			2.61 (6.5)			4.07 (3.7)			2.42 (3.5)	
$\gamma$		0.41 (4.3)			1.01 (3.4)			-0.31 (0.3)			-0.55 (0.7)	
$\kappa$		-0.48 (4.4)				-0.86 (3.7)			1.44 (2.3)			2.37 (2.6)
RSS	381.50	380.02	380.26	213.58	212.69	212.86	328.30	326.89	327.51	172.05	171.08	171.40
Adjusted R <sup>2</sup>	0.985	0.986	0.985	0.990	0.990	0.990	0.939	0.939	0.939	0.938	0.939	0.938

Population, employment, car cost and car time entered as in Table 9

**Table 12** AML Elasticities Implied by Eq. 12

Percentiles	NSS		NSL		SS		SL	
	AML/EGJT	$\eta$	AML/EGJT	$\eta$	AML/EGJT	$\eta$	AML/EGJT	$\eta$
10%	0.034	-0.04	0.008	0.00	0.010	0.00	0.010	-0.01
25%	0.044	-0.06	0.013	-0.01	0.013	0.00	0.013	-0.01
50%	0.061	-0.08	0.019	-0.01	0.018	0.00	0.019	-0.01
75%	0.083	-0.11	0.027	-0.02	0.025	-0.01	0.027	-0.01
90%	0.113	-0.14	0.037	-0.02	0.034	-0.01	0.036	-0.02

Equation 15 provides a simpler model. The inclusion of RGJT compared to Eq. 10 is as intended since there is very little impact on the GJT elasticity. The insignificance of the  $\mu$  term for the NSL model and negative  $\kappa$  not far from -1 is very much consistent with the findings of Eq. 14. Similarly, the NSS model indicates an effect from  $\kappa$  but not proportional, given that the level of AML/RGJT varies around a mean of 0.050 with a 25 and 75 percentiles of 0.031 and 0.059. However, the  $\kappa$  in both season ticket models indicate a relationship the reverse of that implied by the extended GJT approach.

Our conclusions are therefore that it is an empirical issue as to the extent to which the AML elasticity depends upon the proportion that AML forms of GJT. Regardless of the convenience or practice of extended GJT terms, and even with appropriate late time multipliers, it is inappropriate to include AML within an extended GJT without justification and quantification of the implied elasticity variation.

## Conclusions

This paper has presented three related aspects. First, it has reviewed the recent practice of forecasting the impact of late time in the rail industry in Great Britain, as contained in PDFH, set against recent empirical studies that provide important new insights. Second, it has significantly extended the Wardman and Batley (2014) review and meta-analysis of late time elasticities which provided the basis for the prevailing PDFH recommendations. Third, and driven by the review of empirical evidence and identification of gaps in understanding, some fresh econometric insight has been provided into late time elasticities and in particular how they vary. We summarise the findings of each aspect of the paper before providing recommendations for further research.

## Summary of findings

The two most recent econometric studies, OXERA and Winder Phillips (2017) and Steer (2019), challenge current PDFH recommendations, both in implying a lower effect and in demonstrating that a proportional elasticity approach is preferable to the conventional constant elasticity formulation. Indeed, they point to more flexible functions which permit a greater degree of late time elasticity variation. The findings also challenge current recommendations on how long it takes for the long run effect to work through. However, they do not provide entirely consistent evidence and each provides statistically significant demand parameters for less than half of PDFH's flow and ticket type categories.

The meta-analysis presented here was based on a large dataset of 285 late time elasticities assembled from 17 studies, in contrast to the 51 elasticities from 6 studies in Wardman and Batley (2014). It recovers a large number of statistically significant and credible influences. These relate to whether the elasticity was short run, long run or did not distinguish between the two, flow type and distance, season tickets, the measure of lateness, whether the purpose of the study was specifically the estimation of late time elasticities, and whether the elasticity was significant at least at the 10% level. Allowance was made for study quality related issues. The data also seems to indicate that, despite dynamic models being common, there is some uncertainty as to how long the long run is.

An issue relevant to the meta-analysis, and indeed to bespoke econometric analysis, is whether late time parameters that are not significant reflect very low elasticities or whether they are the consequence of the various challenges facing the estimation of late time elasticities. However, the elasticities implied by the meta-model do not vary greatly according to the assumptions made in this regard. We are however of the view that the lower implied elasticities stemming from the use of a parameter relating to the study being explicitly for the purpose of late time elasticity estimation is warranted, given that these studies paid much more attention to the specification of late time variables and the demand data used in the analysis.

The meta-analysis results serve as a useful benchmark and provide some important insights. As with the recent evidence, the meta-model challenges current PDFH recommendations and indicates that lower demand impacts are justified.

The fresh econometric analysis was based on large datasets whose previous analysis form the basis of current PDFH recommendations regarding external factors. The estimated models are robust, obtaining highly significant and plausible parameter estimates for the key influences on rail demand. The analysis is innovative in terms of including late time within an extended GJT term and directly estimating the late time multiplier, and also in terms of specifying and estimating flexible functional forms which explore how late time elasticities vary with the level of late time and GJT.

The econometric analysis also finds that proportional elasticities provide a better explanation of demand variation than constant elasticities and that when more flexible functions are specified the implied elasticity increases more than proportionately with the level of late time. The latter is consistent with travellers not perceiving smaller amounts of late time and/or having non-linear unit valuations of late time, both of which we regard to be plausible but neglected possibilities.

It also finds that if an enhanced GJT approach is to be used, then the late time multipliers used to create it should be directly estimated and are much lower than suggested by prevailing evidence invariably drawn from SP studies. We note that this is in line with the findings drawn by Wardman and Toner (2020) in the broader context of generalised cost and the value of time used to create it. There are obvious implications for modes which cannot directly estimate 'behaviourally consistent' late time valuations. The extent to which the late time elasticity depends upon the level of AML and GJT is one that must be resolved empirically rather than imposed. Finally, in line with other recent evidence, the econometric findings indicate a smaller effect on rail demand than is currently implied in official forecasting guidance.

It strikes us that the above reinforces conclusions that we reached 10 years ago when first publishing in this field, thus: "To date, forecasts of the demand impacts of lateness and reliability have been derived largely from individual-level models taken at a snapshot in time... Whereas individual-level models have suggested a high valuation of lateness and reliability, our market-level models indicate a relatively muted demand response.

Reconciling these findings, we reason that, whilst rail travellers show considerable disdain for experiences of lateness, such experiences will not necessarily dissuade them from travelling by train” (Batley et al. 2011, p 61). Whilst noting the considerable contention that remains in policy and regulatory circles concerning the demand and revenue impacts of rail performance, we feel that our updated and strengthened conclusions largely settle the core argument – improvements in performance will have a limited impact on patronage, and this finding should be reflected within regulatory regimes governing performance and business cases for investment in upgrades to achieve better performance.

Whilst this paper has focussed on the British context, which it is fair to say has pioneered the treatment of how punctuality impacts on rail demand, the elasticity evidence should be transferable to other countries, at least those with comparable rail networks, as at least a first order approximation whilst the methodological insights should also transfer.

## Recommendations

The findings of the various aspects of this paper are consistent in indicating that current PDFH recommendations of the demand impacts of late time variations are too high. Whilst the evidence reviewed and provided here can inform PDFH recommendations (and by implication the regulatory regime, i.e. Schedules 4 and 8), there is clearly a need for further research and the various strands of the paper tend to point in the same direction.

Going forward, there is clearly a need for further research given that the most recent evidence provides statistically robust parameter estimates for only a limited number of key market segments. There is a particular need to focus attention on key markets, such as season tickets, and to be able to distinguish whether insignificant late time elasticity estimates result from very low elasticities or issues arising in estimation. This will require the use of large datasets based on four-weekly data and purpose collected late time data specified at the flow level.

Recent studies and the evidence provided here would seem to suggest that PDFH should move to a proportional elasticity approach and with a somewhat lesser demand impact than currently recommended. However, further analysis must investigate functional form in more detail, rather than simply adopting a directly proportional elasticity function, and allowance should be made for the impacts on late time elasticities of the levels of distance, late time and GJT.

Of particular interest is that the probability that late time is not perceived by travellers is larger for smaller amounts of late time. It might also be the case that marginal utility of late time is not constant. These have implications for how late time is formulated in the demand function. Whilst in principle the estimation process can allow for thresholds and non-linearities in the marginal utility of time and its perception, this might prove a serious challenge given that difficulties can arise in estimating more straightforward functions. Although there have been numerous SP studies of late time valuations, we are not aware that non-linearity of the marginal utility of late time has been investigated. We would therefore recommend that appropriate market research is conducted into both the perception of late time and its marginal utility to inform and enhance the estimation of late time elasticities in econometric demand models.

The evidence here covered indicates that there is some uncertainty as to how long the long run is, even given the widespread estimation of dynamic models. Credible results here are needed, and market research might be able to provide some supportive insights along with theoretical reasoning.



Given that many movements involve interchange, and that interchange can impact significantly on performance, there is a need to extend analysis to cover such movements.

In the longer run, there is a need to determine the impact of the variance in late time on demand, something that has been explored in the past but without great success. Given that the estimation of first order effects are challenging enough, examination of this issue would be wise to avail itself of guiding insights from bespoke market research.

## Appendix: Studies used

The first six studies were used in the Wardman and Batley (2014) review. The rest have been assembled here.

1. ARUP and OXERA (2010) What are the Findings from the Econometric Analysis? Findings Report. Prepared for the Department for Transport, Transport Scotland and Passenger Demand Forecasting Council. <https://www.oxera.com/wp-content/uploads/2018/03/Report-2-Findings-report-1.pdf>.
2. MVA (2008) Econometric Analysis of Long Time Series Rail Passenger Demand Aggregates. Report to Department for Transport.
3. MVA (2009) Regional Flows: Regional Rail Demand Elasticities. Report to Passenger Demand Forecasting Council.
4. Steer Davies Gleave (2003) The Effect of Road Congestion on Rail Demand. Report to Passenger Demand Forecasting Council
5. 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
6. OXERA (2005) How Do Rail Passengers Respond to Change? Report to Passenger Demand Forecasting Council.
7. MVA (2007) London Fare Elasticity Study. Report to the Passenger Demand Forecasting Council, ATOC, London.
8. Steer Davies Gleave (2017) The Impact of Unplanned Disruption on Train Operator Revenue – Peer Review. Report to Passenger Demand Forecasting Council, Rail Delivery Group, London.
9. Steer (2019) Impact of Unplanned Disruption on Train Operator Revenue: Extension of Geography. Report to the Passenger Demand Forecasting Council, Rail Delivery Group, London.
10. Jacobs (2020) Regional Rail Demand. Report to the Passenger Demand Forecasting Council, Rail Delivery Group, London.
11. Jacobs and SYSTRA (2019) Fares Research: Phase 2 Report. Report to Department for Transport.
12. Leigh Fisher, Rand Europe, ITS University of Leeds and SYSTRA (2016) Rail Demand Forecasting Extension. Report to the Department for Transport. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/610059/phase2-rail-demand-forecasting-estimation-study.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/610059/phase2-rail-demand-forecasting-estimation-study.pdf)
13. Meaney, A. and Shepherd, M. (2012) What is Affecting Season Ticket Elasticities in London? Paper presented at European Transport Conference. [https://aetransport.org/public/downloads/\\_KcBy/5392-5218a23dc307c.pdf](https://aetransport.org/public/downloads/_KcBy/5392-5218a23dc307c.pdf)
14. Leigh Fisher (2012) Essex Thameside Econometric Analysis. Report to Department of Transport.

15. Confidential Study for Franchise Bidder (2012)
16. Confidential Study for Train Operating Company (2019)
17. Confidential Study for Franchise Bidder (2012)

**Authors' contributions** Both authors contributed to the study conception, design, funding and manuscript preparation. Data collection and analysis was performed by Mark Wardman. Richard Batley contributed to the review of literature and policy, theoretical issues and to the conclusions. The first draft of the manuscript was largely written by Mark Wardman and reviewed and amended by Richard Batley. Both authors read and approved the final manuscript.

#### Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- ARUP and OXERA.: What are the Findings from the Econometric Analysis? Findings Report. Prepared for the Department for Transport, Transport Scotland and Passenger Demand Forecasting Council. <https://www.oxera.com/wp-content/uploads/2018/03/Report-2-Findings-report-1.pdf> (2010)
- Batley, R.P., Dargay, J.M., Wardman, M.: The impact of lateness and reliability on passenger rail demand. *Transport. Res. E* **47**, 61–72 (2011)
- Button, K.J.: The value and challenges of using meta-analysis in transportation economics. *Transport. Rev.* **39**(3), 293–308 (2019)
- Department for Transport.: Rail Factsheet: 2019. <https://www.gov.uk/government/statistics/rail-factsheet-2019> (2019)
- Elvik, R.: Meta-analytic methods. In: Lord, D., Washington, S. (eds.) *Safe Mobility, Challenges, Methodology and Solutions*, Transport and Sustainability, vol. 11, pp. 423–445. Emerald Publishing, Bingley (2018)
- Gudgeon, L., Martins-Tonks, R., Shepherd, M., Solomon, P.: Do Passengers Respond to Fare Change? Does Inflation Matter? Paper presented at European Transport Conference. <https://aetransport.org/public/downloads/OACOz/4566-55f1b9b79f579.pdf> (2015)
- IBM.: SPSS Statistics for Windows, Version 27.0. IBM Corp, Armonk (2020)
- Leigh Fisher, Rand Europe, ITS University of Leeds and SYSTRA.: Rail Demand Forecasting Extension. Report to the Department for Transport. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/610059/phase2-rail-demand-forecasting-estimation-study.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/610059/phase2-rail-demand-forecasting-estimation-study.pdf) (2016)
- Mott MacDonald and University of Southampton.: Rail Demand and External Impacts: Phase 2 Findings Report. Report to Passenger Demand Forecasting Council, Association of Train Operating Companies, London (2014)
- MVA.: Econometric Analysis of Long Time Series Rail Passenger Demand Aggregates. Report to Department for Transport (2008)
- OXERA and Winder Phillips.: The Impact of Unplanned Disruption on Train Operator Revenue. Report to the Passenger Demand Forecasting Council, Rail Delivery Group, London (2017)

- Rail Delivery Group.: Punctuality Metrics: Punctuality of the National Rail Network. <https://www.raildeliverygroup.com/punctuality.html> (2021)
- Rail Delivery Group.: Passenger Demand Forecasting Handbook: 6th Edition. Passenger Demand Forecasting Council, Rail Delivery Group, London (2018).
- Rail Delivery Group.: Partnership Railway's Transformation in Numbers: Dataset on Rail Industry Finances, Performance and Investment. <https://www.raildeliverygroup.com/files/Publications/2017-12-industry-dataset.pdf> (2017)
- SAS Institute.: SAS version 9.4, Cary, NC, USA (2020)
- Steer Davies Gleave.: The Impact of Unplanned Disruption on Train Operator Revenue—Peer Review. Report to Passenger Demand Forecasting Council, Rail Delivery Group, London (2017)
- Steer.: Impact of Unplanned Disruption on Train Operator Revenue: Extension of Geography. Report to the Passenger Demand Forecasting Council, Rail Delivery Group, London (2019)
- Wardman, M.: Demand for rail travel and the effects of external factors. *Transport. Res. E* **42**(3), 129–148 (2006)
- Wardman, M.: Review and meta-analysis of U.K. Time elasticities of travel demand. *Transportation* **39**, 465–490 (2012)
- 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**, 1041–1069 (2014)
- Wardman, M., Lyons, G.: The digital revolution and worthwhile use of travel time: implications for appraisal and forecasting. *Transportation* **43**(3), 507–530 (2016)
- Wardman, M., Chintakayala, V.P.K., de Jong, G.: Value of travel time in Europe: review and meta-analysis. *Transport. Res. A* **94**, 93–111 (2016)
- Wardman, M., Hatfield, A., Shires, J.D., Ishtaiwi, M.: The sensitivity of rail demand to variations in motor-ing costs: findings from a comparison of methods. *Transport. Res. A* **119**, 181–199 (2019)
- Wardman, M., Toner, J.P.: Is generalised cost justified in travel demand analysis? *Transportation* **47**, 75–108 (2020). <https://doi.org/10.1007/s11116-017-9850-7>
- Wardman, M., Chintakayala, P., Heywood, C.: The valuation and demand impacts of the worthwhile use of travel time with specific reference to the digital revolution and endogeneity. *Transportation* **47**(3), 1515–1540 (2020)
- Wheat, P., Wardman, M.: Effects of timetable related service quality on rail demand. *Transport. Res. A* **95**, 96–108 (2017)

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

**Mark Wardman** is currently an independent researcher with a visiting professorship at the Institute for Transport Studies, University of Leeds. He has been involved in transport research for over 35 years, with a particular emphasis on the analysis of rail travel demand on which subject he has published extensively in international peer-reviewed journals. He has been involved in several econometric studies which have estimated the impact of train service punctuality on demand as well as conducting a number of reviews of late time valuations and elasticities.

**Richard Batley** is Professor of Transport Demand & Valuation at the Institute for Transport Studies, University of Leeds. With a disciplinary background in transport economics, Richard's specialist expertise covers two related areas: first, valuing qualitative aspects of travel (e.g. journey time, punctuality and comfort) in monetary terms, and second, forecasting the impacts of changes in these qualitative aspects on the demand for travel. Richard has operated mainly at the interface between academe and public policy. His work has been reported to senior public servants and politicians, and has been adopted within official policy and practitioner guidance issued to transport operators and scheme promoters.