steer davies gleave

Rail Demand Forecasting Estimation - Peer Review

Department for Transport

Report
2nd November 2016

Our ref: 22983501

steer davies gleave

Rail Demand Forecasting Estimation - Peer Review

Department for Transport

Report
2nd November 2016

Our ref: 22983501

Prepared by:

Steer Davies Gleave
28-32 Upper Ground
London SE1 9PD

+44 20 7910 5000
www.steerdaviesgleave.com

Prepared for:

Department for Transport
Great Minster House
33 Horseferry Road
London

Contents

Introduction

This Peer Review commenced with our receipt of the Emerging Draft Final Report on 28 June 2016. We supplied comments on this report on 20 July 2016. We received responses to these comments on 14 September 2016 and a Draft Final Report on 30 September 2016. This Peer Review is based on the Draft Final Report, with some interpretation as a result of the responses to comments described above. Throughout this Peer Review we recognise that there are limits to the work that could have been done within the constraints of the study, but have noted the areas where in an 'unconstrained' world further improvements could have been made.

This document has been structured in order to address the 8 questions listed in the original ITT. In addition other outstanding issues, that do not readily fall into these 8 categories are noted at the end of this document.

1 Is it possible to understand the methodology used in the study from the final report and associated technical notes?

1.1 We were given access to an emerging draft of the final report and used this to formulate a number of clarification questions. The study team responded to these questions which has helped with our understanding and many of these clarifications have made it through to the draft final version of the report on which this Peer Review is based.

1.2 We note that in some instances, responses have been received as part of this clarification process which do not appear in the final report. For example, the study team suggested that ideally, one would have further data on the production-attraction split for each bi-directional flow in order to more precisely apportion the impact of exogenous variables. It would have been good if this comment, which we agree with, had appeared in the recommendations section, although we are conscious of the practical difficulties involved with this approach.

1.3 The draft final version on which these comments are based is a clear improvement on the emerging draft, although there are still a few weaknesses in this report. For example, there is no indication of document control and no referencing. On page 5, it is stated that the Phase 1 Report should be read as a companion piece to this Final Report. We would have expected the Final Report to be freestanding. We would also have welcomed more signposting and summary of the approach at the beginning of the report.

1.4 Overall, having read through the report a number of times, the approach has largely been understood, although we would have liked to have seen more detail at a number of stages, including worked examples on how the NTS models led to the implied trip rates and onto the use of $INDEX_{SE}$. We understand that there are supplementary files and documents which contain further details, but a review of these is outside of our remit.

2 Are the National Travel Survey (NTS) models built by RAND for the study suitable for the purpose of understanding the propensity of rail trip making by demographic data?

2.1 These models are detailed in Chapter 2 and Annex A. As detailed on page7, NTS involves over 16,000 people per year (300,000/18) but only around 4,000 trips by rail (70,000/18) so that the average number of rail trips per week is only around 0.25, with the vast majority of respondents reporting no trips[1]. Our understanding is that the data used is stage data, meaning that *all* rail trips (including those where rail is not the main mode) will be included.

2.2 Figure 2.1 clearly shows that NTS sampling changed in 2002. Between 1995 and 2001, each annual sample contained approximately 8,000 people and the average number of surface rail trips per year varied between 1,500 and 2,500. Between 2002 and 2012, there were approximately 22,000 people in each annual sample, and the average number of surface rail trips per year had increased to around 5,000. The report does not provide any insight into why and how the NTS methodology changed at this time. If the enhanced sample included a proportionately greater number of people and households with a higher propensity to travel by rail, this could explain some of the increase in the total number of trips between 2001 and 2002. Page 21 notes that the NTS models contain a dummy variable for 2001 to represent the Hatfield rail accident; it might have been useful to see dummy variables for other years (such as 2002 when the sampling methodology changed) in order to attempt to verify the claimed causality.

2.3 Figure 2.2 (page 8) indicates considerable variation over time in average rail trip rates but there is no consideration of standard errors[2] or an attempt to quantify the likely impact of sampling variation. No data are presented for the goodness of fit of the linear trend line in Figure 2.2 – but visual inspection indicates this improves after 2002 – possibly indicating less sampling variation.

2.4 A disaggregate analysis of the NTS is sensible. We assume that number of trips is analysed (in preference to, for example, distance travelled) in order to remain consistent with journeys-based ticket sales regressions. Furthermore, we note that a journey-based dependent variable is also required for the application of a discrete choice model , but note that other modelling forms are possible. The spatial segmentation (London, non-London) is limited (presumably by NTS sample size) and thus cannot exactly replicate PDFH segments.

2.5 The stop/go model is a pragmatic choice as is the adjustment for odd/even combinations. However, the use of the same utility formulation throughout and the lack of a feed-through in

---

[1] From Table 2.1, only 1.8% of persons in the sample make a commuter rail trip, only 0.8% make a business rail trip and only 4.9% make a rail leisure trip per week.

[2] For the 2006 NTS, it was estimated that the mean number of rail trips per year per person was 16.2. The standard error was estimated at 0.76 with a design factor (to take into account the impact of sample weighting) of 1.46 and misspecification factor of 0.98. Source: **https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230547/StandarderrorsMethodology2009NTS.pdf**

terms of a log sum[3] term (or equivalent) could have been considered further. The lack of a logsum measure implies that a Nested Logit approach has not been used; therefore an improvement in the utility of one or all members of the nest will not affect the choice probability of the nest. The appropriateness of the comment concerning simultaneous and sequential choice in footnote 1 (page 9) does not seem relevant – the model is not trying to replicate behavioural processes only outcomes.

2.6 More details should be given of the software used for the estimation (indicated as ALogit in Annex A, page 149) and in particular how the none alternative (0/1+) and stop alternative (stop/go) models are jointly estimated (and note the inconsistency in terminology and with the footnote 1, page 9 comment about the model being sequential).

2.7 As correctly noted by the study team, sensitivity to the choice of the stop number of trips should be considered, but there is some ambiguity as to what has been used in the logit modelling. If we take the number of other trips as an example 95.1% make none, 3.9% make one or two, 0.6% make three or four and only 0.4% make five or more. For one or two stops the probability of stop is 0.8 (3.9/4.9) and for three or four stop it is 0.6 (0.6/1.0). The $p_1 = p_2 = p_3 = \dots p_{stop}$ of the standard model (Annex A, page 113) clearly does not hold. To clarify exactly what was assumed, it might have been useful to include a more thorough exposition of the utility functions for each model.

2.8 On the following page, Annex A clarifies that the standard model is unsuitable in this instance, as it is important to allow for variations of probability at different points in the choice process. We assume that this relates to the inclusion of dummy variables in the stop/go model for specific numbers of trips. This makes theoretical sense, in that both Other and Business trips are relatively well behaved in that the probability of an individual making a given number of trips broadly decreases with the frequency of trips, whereas for commuting the probability of making 10 trips is greater than other categories. Examination of the detailed results on pages 121-126 reveal that, as expected, these dummy variables have a large impact on the probability of making a trip; we feel that these results could have been highlighted in the main text.

2.9 Annex A suggests that a single logit model has been estimated jointly for the none and stop alternative models, which contains <u>all</u> the variables presented in the main text on pages 14-22. We agree that this is the statistically sound way to estimate these models, as opposed to estimating separate models for each 'driver' (such as age, income occupation etc.); the report should have made clear that this was the method used. However, we also assume that these models were estimated separately by geography, which makes it impossible to calculate whether parameter estimates are significantly different between To/From London and Other-Other. Whilst this approach may lead to some estimates which are slightly less robust, it does provide separate estimates for each geographical segment which is consistent with current PDFH practice.

2.10 A feature of the NTS models is that the constants have high statistical significance (as measured by the t-ratios) and have large values relative to the other estimated parameters. As a result the rho-square measure with respect to a constants only model is relatively low,

[3] The logsum parameter, is a function of the underlying correlation between the unobserved components for pairs of alternatives in each nest, and it characterizes the degree of substitutability between those alternatives. It can be used to create a combined logsum utility for these alternatives.

although the rho-square measure with respect to a null model is high. Because the models are jointly estimated it is not possible to assess the goodness of fit of the none alternative and stop alternative models separately although it is suspected that the former has the superior fit. (From Tables A.7 to A.9. It is not clear what ^ refers to in these Tables).

2.11 The NTS data used is pooled cross-sectional and longitudinal (pseudo-panel). The econometric issues that might arise are not considered. The longitudinal nature is dealt with by a linear time trend (although for business this is limited to after 2006) and the dummying out of some years (2001 relates to Hatfield, but it is not self-evident why 1999 is relevant for business – the subsequent company car tax explanation (page 22) has not been fully explained, but is likely to draw from other studies which have assumed a similar effect). There is no attempt to relate the trends here with subsequent analysis of the GJT time trend.

2.12 More details could have been supplied of the computation of the calibration factor and why the odd/even fraction is held constant across all alternatives (page 22). The extent to which the calibrated models replicate Table 2.1 should be assessed. It appears that the dummy variables used for different numbers of trips (1-2 trips for business/other and 9-10 trips for commuting) already take account of this difference. Again, it might have been advantageous to include a broader narrative as part of Annex A.

2.13 No consideration is given of alternatives to the stop go model, such as a linear regression. However, as mentioned in paragraph 2.4 above, this model specification works well given that the dependent variable is measured in numbers of journeys. An alternative specification would therefore have been more applicable if the unit of measurement was passenger miles, but this is likely to have been outside the scope of the project.

2.14 When examining the results of the NTS regressions in the main report, care is needed when interpreting dummy variables for economic status and occupation. If all occupations are classified as either full time or part time employment, then it is unclear how all these variables can appear in a single regression without perfect multicollinearity occurring. This applies particularly to Table 2.7, where ‘Admin. Occupation’ is described as being part of the base, but is associated with a non-zero coefficient. This specific case aside, the detailed results in Annex A do provide a clear demonstration of how such variables are treated.

2.15 **Although there are some minor concerns over some elements of the NTS modelling work, these are largely presentational. Given the relative paucity of work in this area for rail (with one or two exceptions), this is likely to represent an advance.**

3 Is the construction of the two new variables (known in the study and technical documents as “POP_INDEX_{SE}” and “EMP_INDEX_{SE}”) that have been used in the regression of ticket sales data, and are based on the results of the NTS model, fit for purpose?

3.1 The NTS is used to determine expected trip rates as shown by Tables 3.2 to 3.4 (where it refers to implied trip rates). How this is done is not fully explained. Also, the translation from NTS to RUDD is not clear. For example, for the commuting model the age bands in the RUDD based trip rates (Table 3.2) are different to those in the equivalent NTS commuting model (Table A.7).

3.2 Furthermore, the NTS models include more variables than are listed in Tables 3.2 to 3.4, including a time trend variable. After reviewing the Emerging Draft Final Report, we asked a question to clarify our understanding that coefficient estimates do not vary over time. However, the answer did not permit us to better understand what is assumed for the variables not listed in Tables 3.2 to 3.4 which appear in the NTS models. We now understand that to calculate the deviations, the logit regressions are ‘re-run’ for each segment (e.g. only containing respondents aged 15-29) and then the coefficient estimates from the remaining variables used to calculate a revised trip rate. In order to aid the reader’s understanding, it would have been good to have seen a worked example of this for a selection of deviations.

3.3 In addition, some discussion is required of the variables that are included in the NTS models but not included in the RUDD models and the implications of dropping the licence holding and company car variables for Other trips to/from London – some more details are given in Table A.12 which show these parameter estimates are statistically significant but of the wrong sign. This might indicate multicollinearity. The recommended values given in Tables 7.2 and 7.3 are clearly derived from Tables 3.2 to 3.4, which is slightly obscured by the re-ordering of the columns but with the car licence element dropped. Ideally, this would have involved recalibration of the NTS models.

3.4 The treatment of long distance as bidirectional is presented throughout the report as an improvement on the methodology. While it does mean that the weakness of the data as regards the production and attraction stations for single leg (usually Advance) fares is recognised, alternative assumptions have to be made and we have some concerns that these are equally as flawed and are applied inconsistently throughout the study. Table 3.6 indicates that populations are averaged but it is not clear if this is a weighted average; a straight average would appear to be as flawed as assuming that the flow is unidirectional. We also understand that this is inconsistent with how bi-directional flows are treated in the ticket sales models where we understand that the relative size of the stations is used to create a 100% producer – attractor flow. Our belief is that, on the vast majority of flows, where Advance is a relatively small part of the market, standard uni-directional models would be superior, and for the remainder ticket fulfilment analysis could yield a reasonable proxy for the correct producer attractor split – although we do accept that this additional work would have been outside the scope of this study.

3.5 For Non London Short Non Season and the Network Area to London Non Seasons markets, both population and employment variables are examined. It is also possible that business trip making on Long Distance flows may be determined by employment as well as population – but this does not seem to have been tested.

3.6 A modification is explored in which the indices are adjusted for income ($INDEX_{SE\_INC}$) but this does not lead to improvements compared to $INDEX_{SE}$. The income elasticity (equation give on page 15) is an arc elasticity (which will vary) in contrast to the constant point elasticities used in PDFH. There are very large variations between London and Other (Table 2.4). It is also not fully explored why $INDEX_{SE\_INC}$ is not taken forward; page 45 states that there are "concerns that it might detect temporal income effects otherwise attributable to GVA". Whilst we see this as a valid concern, such effects could have been explored further in the context of the ticket sales models.

3.7 When the $INDEX_{SE}$ variable is tested for goodness of fit (Table 3.13), it minimised the residual sum of squares (RSS) for four of the six flow types. However, for arguably the two most important flows (London Long and Network Area to London seasons) the PDFH model gives the best fit. Although these two flow categories only account for 7% of observations, they are likely to account for a much bigger percentage of volume. Similarly, in Section 4, the best model as determined by RSS does not include an $INDEX_{SE}$ measure for Long Distance London Flows (Table 4.3) and Network Area to London Seasons (Table 4.10)

3.8 In summary, the principle of improved demographics is a good one, as this is an area which is largely neglected in current rail demand forecasting practice. Although in some cases the POP_INDEX_{SE} and EMP_INDEX_{SE}, as applied in practice do not lead to a material improvement as measured by goodness of fit, the theory supporting these variables is sound. Furthermore, the NTS models underpinning POP_INDEX_{SE} and EMP_INDEX_{SE} are generally sound. In order to fully understand how these NTS results are used to calculate the revised indices, we believe that the main report would have benefited from a fully worked example as part of the main text.

4 Does the variable selection and the specification of the ticket sales regressions used in the study make theoretical sense?

4.1 The general principles of the modelling approach are outlined in Section 3.4. This is based on a fixed effects regression, estimated using Ordinary Least Squares with a pooled data set of routes and years. Pooling data has an advantage of increasing the number of observations, with resultant very large data sets, particularly for non-London flows[4]. However, pooling data leads to statistical problems related to heterogeneity bias. This is dealt with in the fixed effects model by introducing route specific dummy variables (this approach can also be replicated by using a semi-automated method such as mean deviated regression).

4.2 Although recognising that this is not primarily an exercise in econometrics, one would have normally expected the Fixed Effects (FE) specification to be compared with a Random Effects (RE)[5] specification and with Pooled Ordinary Least Squares (POLS)[6]. A series of diagnostic test would then be run such as the incremental F-test (to compare FE and POLS), a Hausman Chi-Square test (to compare FE and RE) and a Breusch-Pagan Lagrange test (to compare POLS and RE). It is our experience with data of this type that the FE specification provides the best fit but for this work we need to take this on-trust. This should be an area of further work, particularly if the preferred models are taken forward.

4.3 It is stated on page 37 that 90% of the fixed effects dummy variable coefficient estimates are significant at the 10% level; it is unclear why 5% is not used. However, it is clear that these dummy variables are providing most of the explanatory power, as could be expected given the amount of variation between different routes. From Table 4.3, it can be seen that the fixed effects only model has an adjusted R^2 of 0.970, whilst for the preferred model this is 0.989. This implies that the explanatory variables in the preferred model are explaining around two thirds of the 'remaining' variation.

4.4 On page 38 it is stated that "The t ratios for key variables are generally so high that correction for the presence of autocorrelation or heteroscedasticity would not materially alter the confidence we can place in the parameter estimates". Although this statement may be true in most practical applications, the presence of heteroscedasticity or autocorrelation can mean that the parameter estimates are themselves unreliable. In this application, this would be most likely to occur due to the omission of (potentially explanatory) lagged variables; if this was the case, then it is possible that certain parameter estimates may exhibit bias.

4.5 We would normally expect to see some testing for multicollinearity (such as variance inflation factors)[7], heteroscedasticity (such as the Likelihood Ratio and Wald tests for panel data) and

---

[4] 9,712 for Long Distance London, 117,296 for Long Distance Non London, 10,604 for Network Area London Non Seasons, 8,075 for Network Area London Seasons, 69,319 for Non-London Short Non-seasons, 34,786 for Non London Seasons

[5] Readily available in statistical packages such as Stata and can be replicated with PROC GLM in SAS.

[6] Although one would expect POLS in this instance to be plagued by ecological fallacy issues – which may not be totally eradicated with a fixed effects model.

[7] Some correlation coefficients are provided, for example, between the time trend and other variables in Table 4.0.

autocorrelation (such as the Wooldridge test for panel data[8]), along with appropriate references to the literature. Methods such as Heteroscedasticity-consistent standard error estimators might be considered. On page 38, there is some indication that weighted least squares (WLS) was examined, presumably to deal with heteroscedasticity, but few details are given. Outliers have been dealt with by removing observations with standardized residuals greater than plus or minus 2.[9] This might be seen as a form of WLS in that these observations are given zero weighting. Multicollinearity is apparent given the difficulties that arise from free estimation of parameter values and the resultant need to constrain estimates to PDFH values and this is recognised in the text (e.g. page 56).

4.6 The general specification of the model estimated is given by the equation on page 37. This is a multiplicative function that can be linearised through a logarithmic transform to give (at least for the continuous variables) constant elasticities that are consistent with the form used in the Passenger Demand Forecasting Handbook (PDFH). The appropriateness of alternative functional forms, such as the negative exponential, has not been tested, although Unemployment appears to enter in this form.

4.7 Although not explicitly discussed, the variables selected are those that one would expect in a standard transport demand model, including own price and journey time, the price and journey time of rival modes and various variables to capture taste variation, including income, population characteristics and employment characteristics. The variable selection is, in any event, largely constrained by what appears in the PDFH. Ideally the write-up in Section 4 would have been re-ordered as the initial discussion is of variables (such as reliability, unemployment, rolling stock and gating) and of interaction effects which, in the main, do not appear in the preferred models. A potential omission is consideration of competition between routes/stations and the impact this may have on station catchment areas. This is a major weakness of PDFH which is not addressed in this study.

4.8 It is claimed on page 42 that the elasticities freely estimated with this model are dealing with a one year response[10], whereas the PDFH elasticities are explicitly long run (with PDFH5.1 indicating responses taking up to five years from major commuting service changes, for example). One might have expected some exploration of dynamic panel effect models, either in levels or differences, although it is accepted that past attempts at this have not been particularly successful. This is highlighted as an area for further work. We further acknowledge that in general, elasticities currently recommended by PDFH have not been derived from dynamic models.

4.9 As discussed in pages 38-43, a number of parameters have been constrained in the ticket sales regressions. In a number of number of cases these constraints are likely to be necessary; however, the constraining of POP_INDEX_{SE} and EMP_INDEX_{SE} to one in all preferred models appears to be on the basis of a prior-held belief rather than statistical evidence. Crucially, it ensures that any differences between Population/Employment and their 'socioeconomic'

---

[8] Rather than Durbin Watson – which has been used (page 38)

[9] For London Long Distance, there is a loss of 4.8% of the data points for an increase in the adjusted R-square from 0.978 to 0.989 (Table 4.3),For Non London Seasons there is a loss of 4.5% of data points for an increase in the adjusted R-square from 0.753 to 0.864 (Table 4.14).

[10] Although it could be argued that, if the market is not in equilibrium, demand in year t is affected by explanatory variables in previous years.

counterparts are always assumed to have a demand impact. Although there is certainly a case for constraining these parameters (given the potential for nonsensical results if they were allowed to vary freely) it would have been interesting to see a repeat of the test results in Table 3.13 where POP_INDEX_{SE} is allowed to vary freely. In some cases, whilst the 'base' parameter is constrained to one, multiplicative dummies are estimated which allow significant variation, which require further judgement. For example, a distance effect could have been applied to the GVA elasticity in the London Long Distance segment (as for the Non-London segments), but was ruled out because the effect seemed too large.

4.10 **In summary, the general specification and variable selection of the ticket sales model follows PDFH practice and is broadly consistent with micro-economic theory, although the lack of consideration of lags is a weakness. The specification in terms of fixed effects may be appropriate – it certainly results in good fit as measured by the adjusted R^2 (ranging from 0.843[11] to 0.989) of the preferred models and high t-ratios. However, some further statistical tests would ideally have been undertaken to confirm that this specification is appropriate.**

[11] The fit for the non-London seasons model (Table 4.14) is notably worse than the other models, but this is likely to be because there is less cross-sectional variation between routes and hence the fixed effect dummy variables have less explanatory power.

5 Do the recommendations of this study make sense given the results of the ticket sales regressions and the NTS models?

5.1 The preferred models are given by column I of Tables 4.3, 4.5, 4.8, 4.10, 4.12 and 4.14. In none of these cases is the preferred model the one that gives the best fit – so the choice relies on an element of judgement.

5.2 The GJT trend variable appears in most of the preferred models - although not for Network Area to London. Its basis is described on page 50. But this seems somewhat tenuous, based on the finding of Wardman et al. (2015)[12] that the value of travel time for train was 69% of the gross wage rate in 1986, falling to 32% in 2009 - equivalent to a 3.3% per annum reduction in the value, which will be offset by the growth in real wages. This may be the basis for the 1% reduction in GJT per annum, except for commuting to London where this is assumed to be dissipated by overcrowding. This decline is related to the use of mobile technology with an assumed start date of 2000. It would be more convincing if this trend was related to sales of digital technology devices or some related measure (e.g. of Wi-Fi connectivity); the report does mention that this was attempted. It is also worth noting that the way the GJT trend is specified (as a fixed 1% point improvement on GJT per annum), its impact increases over time.

5.3 It is unrealistic to assume that this trend will continue indefinitely – digital technology will eventually reach saturation point. While advances in mobile technology have been apparent over this time period, so have a number of operator investments across the privatised railway, that a 1% per annum improvement in GJT could also describe. Inclusion of this variable does undoubtedly improve the fit of the models (both new and PDFH), although clear guidance is given that this variable should not just be rolled forward into the forecast (for the reasons described above), but should still be considered.

5.4 For London Long Distance some dummies related to Hull Trains, Grand Central and West Coast disruption are significant. For non-London Long Distance, of the dummies only the West Coast disruption variable remains in the preferred model. An additional GVA effect is detected, associated with 11 Core Cities, which seems plausible.

5.5 For the Network Area – London Non-Seasons model a TKT_Index variable is included to take into account of switching by some commuters from season tickets. The basis for this is given by Figure 4.1 (although it is not clear whether this figure refers to all rail commuting or Network Area commuting to/from London) and the equation on page 61. Goodness of fit and statistical significance measures are not provided and it is also evident from the figure that this effect is only examined from 2002 onwards, although the text on page 62 refers to 1995/6.

5.6 For London and the Network Area non seasons, it was found that the GVA elasticity for flows from London was lower, which is not unsurprising.

5.7 For the Network Area to London seasons, the Unem(ployment)[13] parameter value is positive which at first seems counter-intuitive. It is only intuitive if it is assumed that the performance of the labour market at the destination (central London) is much stronger than the origin. This clearly has been the case over the period of study but may not hold in the future (for example,

---

[12] It would have been good to include a reference to the relevant section in the report.

[13] The definition of this variable in Table 4.1 seems to suggest that this is a participation rate and there is some hint of this in the main text, which would explain the discrepancy here.

if there was an attempt to re-balance growth in the South East). It is possible that a different variable, such as the net balance between jobs and workers for both the origin and destination locations, might be more intuitive theoretically. Large flows would be expected from origins with a surplus of population to destinations with a surplus of jobs.

5.8 For Non London short non seasons, the fare elasticity and GVA elasticity for PTE areas are lower (in absolute terms), which seems plausible for fares, but less so for GVA. However, there are also positive increments related to GVA_NCM (Neither to Core/Major[14]) and GVA_CMN (Core/Major to Neither). There is also a positive increment for $EMP\_INDEX_{SE\_PTE}$.

5.9 For Non London seasons, there seems to be a fare elasticity increment for long distance flows (20 to 50 miles) which seems plausible. Table 5.5.1 (Non-London Seasons) contains a fares parameter of -0.99 for the preferred model >20 miles, which is obtained as a combination of the Fare_Short and Fare_Long parameters in table 4.14. Therefore, it should be clarified that the _Short variables represent the 'base' and that the _Long variable parameters should be added to those for the _Short (this is standard practice when dealing with multiplicative dummy variables).

5.10 There are also EMP_INDEX_{SE} increments for the core cities and longer distance flows. A positive unemployment coefficient value is again used, presumably in this case reflecting the stronger performance of the labour market in the core cities over the study period. This model has modest goodness of fit and Table 4.15 illustrates very mixed growth rates for the 11 anonymised core cities, in part reflecting the economic downturn from 2007 onwards but also differential patterns (spatial, temporal and structural) of economic restructuring.

5.11 The recommended elasticities that emerge from these models are discussed from page 79 onwards. The wording of the additional component to GJT is slightly unclear, but the overall message is understood. The unemployment rates seems to have been reinterpreted as a participation rate (with a sign change). It is also evident that the unemployment is in fact the participation rate and is entered in unlogged form (see also discussion on page 65).

5.12 **In conclusion, many of the results are sensible in a practical sense, which was (with some justification) the main criterion for selection. A possible exception is the unemployment rate which in the recommendations is being reinterpreted as a participation rate. Whilst it seems sensible to us that unemployment (or at least differences in employment) should be allowed to influence rail demand, it is possible that this variable could be replaced with a more theoretically justifiable variable. Given the way in which the preferred models were selected, the statistical justification of some choices is weak; this may be partially reflected by the relatively modest performance of some models in the backcasting exercise**.

---

[14] The definition of major might need reviewing. For example, Cambridge is included, Oxford is not. Norwich and Ipswich are included, but Portsmouth and Southampton are not.

6 Does the peer reviewer have any other thoughts on what is positive and negative in the RDFE study and it its methodology?

6.1 Positives:

- Use of NTS is innovative and gives valuable information on the underlying 'propensity' of certain socio-economic demographics to use rail.
- Some useful additions have been made to the RUDD dataset, and in general the study seems to have provided good value for the available budget.
- GVA elasticities seem sensible on the whole, and show signs that some variation which was previously explained by economic growth may actually be due to shifts in population demographics.
- The study has undergone an internal quality assurance process, which lends weight to the robustness of the results that are presented.
- The large NTS dataset, which has been used to enhance RUDD, again improves the degree of confidence in the estimated parameters.
- Furthermore, the data (RUDD and NTS) is considerably more up to date than the vast majority of data used to generate values used in PDFH.
- Use of a ticket switching index seems sensible and does tie in with well observed trends in passenger behaviour.
- Econometric models used and coefficients obtained *appear* to be generally sound from a statistical point of view and generally improve the back-cast.

6.2 Negatives

- There is a lack of statistical testing in general, both in the NTS logit models and the tickets sales regressions. The choice of preferred model also is to driven by judgement rather than by statistical evidence. Whilst there is a potentially strong justification for this, there is the possibility that the study outcomes could continue to reflect a long-standing weakness in the PDFH approach.
- The explanation of how the POP_INDEX_{SE} and EMP_INDEX_{SE} are created, stage by stage, from the NTS logit regressions is at times unclear and does not enable the reader to fully understand the steps that have been taken. We feel that this explanation would have merited a place in the main text, given the importance of the new variables.
- The declining GJT elasticity has a **lot** of explanatory power. It is based on a fairly crude assumption (although the idea of trying to adjust for unobserved endogenous improvements to rail journeys is a good one) and leaves the question as to how to factor in technological and endogenous changes going forward, given that the study does not recommend using this parameter for forecasting purposes.

7 Does the peer reviewer have any thoughts on how the study's findings should be best implemented in rail demand forecasts?

7.1 There are two main themes surrounding the study's applicability:

- The ongoing use of the technology time trend and other terms used in the preferred models; and
- The routine use of POP_INDEX_{SE} and EMP_INDEX_{SE} variables, which may present some practical challenges

Terms outside of the framework

7.2 Section 6.3.6 sets out the study team's observations about the use of the GJT_Trend term. This section notes that the inclusion of this term in the preferred models has reduced the recommended elasticities to GVA, but, at the same time, guards against the blind continuing use of such a time trend, which would continue to assume improvements in technology having a similar impact. Instead it recommends ensuring that continuing technological development is one of many impacts, most of which are endogenous to the TOC, that should be considered on top of the new elasticity recommendations – as will already be the case in many applications.

7.3 We do not consider that the actual term that has been included in the models has strong evidence behind it and so it is difficult to conclude whether or not it does continue. It is worth noting that recent work that we have undertaken for PDFC suggests that in the recent past revenue growth has outperformed existing PDFH guidance, but that it has reverted closer to what PDFH would suggest over the last year. Although this is shorter term trend, it would appear to be consistent with the suggestion that this time trend should be excluded, or at least reduced, going forward.

7.4 There are other terms that were included in the preferred models, but have not made it through to the study recommendations, these include:

- Disruption (West Coast in this case);
- Ticket switching index;
- Using CPI to deflate fares changes.

7.5 Clearly the first of these factors is unique to the period of the analysis; it is also plausible that the other two would not be expected to be carried forward, although this is less clear-cut.

7.6 In practice TOCs will include factors such as station and rolling stock quality, benefits in retailing, WiFi etc in their forecasting and will continue to do so. If this study's recommendations are to replace those in PDFH, at a high level the only change they will see will be a reduction in the elasticity to GVA and the impact of the POP_INDEX_{SE} and EMP_INDEX_{SE}. Therefore, the net effects of these changes is likely to be a reduction in their forecasts which may be difficult for users to accept.

POP_INDEX_{SE} and EMP_INDEX_{SE}

7.7 As Section 6.5 notes "Implementing the recommended forecasting framework will make rail forecasting more complicated than the traditional Passenger Demand Forecasting Handbook approach". This is largely because of the need to calculate and apply the $INDEX_{SE}$ variables. It is also true to say that forecasts for the required socio-economic data are not routinely available at a high level of dis-aggregation, although data is available at a regional level.

7.8 We have already made the suggestion that worked examples should be included to make the report more auditable. Furthermore, we suggest that this should be extended to the development of spreadsheet models to facilitate their calculation for station pairs, purpose and segment combinations. In addition it seems clear that DfT should be involved (at least initially) in the development of appropriately disaggregated socio-economic variables. We understand that this work has been started.

7.9 We would suggest that the next step is the development of a series of “road testing” forecasts using the recommendations and any newly sourced socio-economic variables for different applications of the recommendations in consultation with practitioners at TOCs, owning groups, Government and consultants. This should include forecasts which would routinely include the sorts of factors discussed in Section 6.3.6, in order that we might understand the impact of the changes on forecasts that already include factors over and above the core PDFH external factors. We understand that work on road testing has been started by DfT.

7.10 It is also worth noting that should the new recommendations be adopted, the complexity of the recommendations is likely to lead to short-cuts being developed for high level use by TOCs and other users. This may extend to either applying the POP_INDEX_{SE} and EMP_INDEX_{SE} elasticities to unindexed population and employment, or, more robustly, developing TOC/route specific rules of thumb for overall average weights to apply. High level use of the recommendations is inevitable and some work should be undertaken to understand the likely impact on forecasts.

8 Given the answers to the previous questions, and given the fact that the data used for this study is more up to date than all of the evidence used in studies in PDFH, can the peer reviewer conclude whether the recommendations of this study are more fit for purpose than existing WebTAG/PDFH evidence?

8.1 There are undoubtedly advantages of the current study over the previous studies on which the PDFH guidance is based. The study uses up to date data, which has been quality assured and developed by a number of independent industry experts over recent years. The study itself has been subject to levels of Quality Assurance not seen in previous studies, not least by this Peer Review. The study has also produced consistent results across the different geographical segments, although the performance of the recommendations across the different segments does vary markedly.

8.2 We recognise the limitations of the study, in terms of budget and timescales, has meant that it has not been possible to test all possible approaches at all stages on the analysis. We also recognise that judgement has had to be applied to the results from the statistical analysis and that the recommendations are based on a balance of statistical significance, performance in the backcasting and judgement.

8.3 It is clear that the modelling of propensity to use rail using NTS data is innovative and brings additional information to the approach. We have concerns that the NTS data on which this is based does vary over time and that consideration should be given to updating the resulting indices for future application.

8.4 We do have concerns about the constraining of parameters to current PDFH values, but we recognise that this approach is likely to return more usable recommendations. We are aware that recent studies which have given freer rein to estimate parameters have resulted in implausible recommendations which are not applicable for forecasting and recognise that the study has sought to demonstrate (albeit in a limited way) that this could happen in this case.

8.5 Despite the logic behind constraining parameters, judgement has been applied which has resulted in the (partially) free estimation of otherwise constrained parameters for some segments, most notably Non London long distance and core city employment elasticities. We recognise that a special case has been made in these segments where the increased parameter is in effect compensating for inferior data capture in non-London city centres. A more robust health warning should be applied to these elasticities in the scenario where future data capture is improved, or a substantial overestimate of the impact of this driver could result.

8.6 Despite our concerns about constraining of parameters, it is a key strength of the study that the current approach is able to support guidance which is broadly consistent with existing evidence using an independent approach and up-to-date data.

8.7 We have mixed views on the benefit of the inclusion of the additional explanatory variables in the ticket sales analysis and the backcasting – predominantly the GJT time trend and ticket switching index. It seems clear that these variables do improve the historical fit, but there is a concern that their impact is unknown going forward. On the one hand the demonstration that the inclusion of these, albeit rather blunt, variables does improve the model fit shows that the modelling framework can be applied going forward in the way that PDFH is routinely applied, with endogenous drivers over and above PDFH exogenous drivers. However, forecast data on the specific drivers used is not available and as endogenous drivers are routinely applied on top of existing exogenous guidance, the impact is likely to be a simple reduction in forecasts going forward. While this does fit with the current slowdown in revenue growth, this appears to be coincidental.

8.8 While we are sympathetic to the suggestions concerning bus competition, we do not consider that the same level of innovation and thorough analysis has been undertaken on these elasticities. Therefore, while it would be preferable to support the study's recommendations wholesale, we do not support these recommendation and believe that further work should be undertaken to update the PDFH guidance on bus competition.

8.9 Considering all of the above we conclude that the study recommendations offer an improvement to the existing PDFH guidance which is an amalgam of evidence gathered largely from the NPDFF study of over 15 years ago and more ad hoc more recent studies, particularly outside London. There are some provisos to this conclusion as described above, and we do feel that the case for adoption of the recommendations of the study by PDFH would be better made with improved documentation and particularly testing – both of which we understand are taking place.

9 Additional Comments

9.1 There is a general requirement for a greater degree of explanation, particularly around the creation of the POP_INDEX_{SE} and EMP_INDEX_{SE} variables. In particular, it would significantly improve our understanding if the following had been provided:

- A worked example, for a given journey purpose, geography and segment, how the implied trip rate is calculated from the relevant coefficients in the NTS models. This would include an exposition of the utility functions for the NTS models, as well as an explanation of how the odd/even fraction is applied;
- A worked example, for a given flow type, how the expected trip rate is calculated and hence how $INDEX_{SE}$ is calculated.

9.2 With respect to the backcasting, if the quantities in the graphs are added up for 2015, the actuals come to around 23.5 billion passenger miles. ORR reports that total usage (excluding Heathrow Express, Eurostar etc.) was around 39.6 billion passenger miles in 2015. In other words, only around 59% of total demand is modelled. Contrary to page 25, this is unlikely to be 'the vast majority of rail revenue in Great Britain.' Some discussion of the grossing-up factors over time would have been be useful, particularly given that the data is truncated (RUDD flows of at least £10k nominal in 2005/6 – this precludes new stations and services opened after March 2006, for example).

9.3 If we base the accuracy of the backcasting models on the unweighted means of the absolute deviations of the CAGR in the Tables from page 81 onwards (and these Tables, and others in Chapter 5, should be numbered), the preferred model provides the best fit for Long Distance London (Ordinary), Network Area to/from London (Ordinary) and Network Area to London (Seasons), the latter jointly with the preferred model with PDFH4. For non-London flows, the preferred models with PDFH4 fares seem to be preferable.

9.4 The recommendations concerning bus and coach competition (page 105 onwards) only hold if the bus own price elasticity and the diversion rate are held constant. Over the (very) long term there is evidence that bus price elasticity has been increasing (see, for example, the Demand for Public Transport: a Practical Guide) and, given rail's growth, it is also likely that the proportion of bus growth (or decline) that is abstracted from (or diverted to) rail will increase. In combination, this could offset the change in market shares – where the bus market is getting smaller relative to the rail market. Hence the basis for changing these elasticities seems weak. Furthermore, in Table 6.8, it is not clear that the national surface rail passenger miles refer to non-London instances.

9.5 Goodness of fit statistics for the modelling of car and bus speeds (page 132) are only modest (R2 between 0.31 and 0.37) suggesting some of the unexplained variation in the models will be due to measurement error rather than specification error.