OFFICE OF RAIL AND ROAD

PR18 Econometric top-down benchmarking of Network Rail

A report

July 2018

Contents

PR18 Econometric top-down benchmarking of Network Rail

Context

1. The 2018 periodic review (PR18) is the process through which we determine what Network Rail should deliver in respect of its role in operating, maintaining and renewing its network in control period 6 (CP6)[1] and how the funding available should be best used to achieve this. One of the principal purposes of PR18, as set out in our initial consultation document, was to establish a more efficient and better-used railway, delivering value for passengers, freight customers and taxpayers in CP6 and beyond.

2. Scrutinising Network Rail's cost proposals and delivery planning is key to this overall objective. The PR18 efficient cost work stream was established to undertake this scrutiny with the ultimate aim of setting challenging but achievable efficiency savings targets for Network Rail during the next control period (CP6). As part of the efficient cost analysis, we are using different approaches including econometric top-down benchmarking that we present in this technical paper.

3. This paper summarises the findings of the two strands of econometric top-down benchmarking that we have undertaken in PR18.These are

   - Econometric top-down benchmarking of Network Rail total maintenance and renewal costs; and
   - Econometric top-down benchmarking of Network Rail's maintenance delivery units' maintenance costs.

4. This analysis' findings supported our understanding of costs and helped us to identify where there might be evidence of performance issues. However, the direct use of these findings in making our PR18 decision on Network Rail's efficiency targets is limited, because their robustness is constrained by data quantity and quality. Nevertheless, our econometric benchmarking work indicates that there are potential inefficiencies within Network Rail. This supports our findings in other areas, notably the review of headwinds and efficiencies as described in our PR18 draft determination (see the overview here).

5. In our PR18 draft determination, we described how we envisage to make changes to how we regulate Network Rail, with a greater focus on route-level regulation and a

[1] CP6 will run from 1 April 2019 to 31 March 2024

targeted approach to regulating the System Operator (SO) in CP6. We emphasised that our CP6 regulation of Network Rail will make greater use of comparison between routes. We expect this approach to sharpen the incentives on each route to perform and to provide a stimulus to sharing of best practices across Network Rail.

6. Compared to previous periodic reviews (PR08 and PR13) where we used international data to benchmark Network Rail against other similar infrastructure managers in Europe, this emphasis on intra-routes comparison in CP6 and beyond constitutes an opportunity that we will take to make more use of top-down econometric benchmarking.

7. Although this analysis was constrained by data quality and quantity as it is the first time that we conduct econometric benchmarking at such a geographically disaggregated level, this analysis constitutes the foundation for our future intra-Network Rail econometric benchmarking.

Purpose and structure

8. This paper summarises the findings of the two technical reports of the econometric top-down benchmarking work that we have undertaken as part of PR18 efficient cost analysis. These two technical reports constitute one element of the overall evidence base for our assessment of Network Rail's proposed cost and delivery plans. More details on the methodology, data and findings can be found in the attached annexes A and B. Annex A discusses the econometric top-down benchmarking of Network Rail routes while annex B discusses the econometric top-down benchmarking of Network Rail's maintenance delivery units (MDUs).

9. This summary paper is structured as follows: in addition to the general context presented above, section 1 presents the objectives of our PR18 top-down econometric benchmarking analysis, section 2 presents the methodologies we used, section 3 summarises the findings while section 4 concludes.

1. Objectives of the analysis

10. The overall objective of the PR18 efficient cost work stream is to scrutinise Network Rail's cost proposals and delivery planning to ensure that they are stretching enough but achievable. To deliver on this objective, one of the approaches we have adopted is to use econometric top-down benchmarking to undertake intra-Network Rail cost benchmarking.

11. This has involved using econometrics to benchmark Network Rail's routes total maintenance and renewals costs on the one hand and maintenance delivery units'

maintenance costs (MDUs)[2] on the other. Using historical data, this analysis pointed to potential inefficiencies in Network Rail's maintenance and renewals planning and delivery processes. This supports our findings in other areas, notably the review of efficiencies. This analysis was constrained by data quality and quantity, but it established a basis for improvements to data quality and a way forward for benchmarking initiatives to inform our ongoing regulatory activities.

12. In previous periodic reviews (PR08 and PR13) we used international data to benchmark Network Rail's performance against similar infrastructure managers in Europe. In comparison, the present intra-Network Rail benchmarking gives us insightful information about performance within Network Rail. It offers some technical advantages, in particular in the consistency of data definitions. In addition, while there is still obvious heterogeneity among Network Rail's routes and MDUs that our analysis has not controlled for, the assumption that they operate in broadly similar conditions (which means they are relatively easy to compare) is more acceptable than making the same assumption about Network Rail vis a vis its European peers.

13. We had planned to undertake international benchmarking similar to previous periodic reviews using the Lasting Infrastructure Cost Benchmarking (LICB) data to benchmark Network Rail against its European peers. Network Rail is a member of the LICB project and had previously provided us with data, but technical and administrative barriers meant that it was not able to do so in time for the present PR18 analysis.

14. In carrying out this analysis, we worked closely with Network Rail, who not only provided us with the necessary data but also worked constructively with us to cleanse it and correct errors and inconsistencies that we identified.

2. Methodology

Introduction

15. In this section, we briefly discuss our approach to the benchmarking analysis. More details are available in the two papers in Appendices A and B. In this analysis, we used statistical techniques that are widely used by regulators (to analyse and challenge regulated companies' efficiency) and in academic research (to undertake cost frontier analysis). We broadly refer to these statistical techniques as econometric top-down benchmarking.

16. The standard approach to econometric top-down benchmarking is to use historical data to estimate a cost function. This is then used to produce efficiency scores for

[2] MDUs are operating units within routes that are responsible for most of Network Rail's maintenance activity

each comparator, typically relative to the most efficient peer(s). The econometric approach simultaneously takes account of variation in several cost drivers. This means that econometric top-down benchmarking can be used to estimate the impact on costs of different variables – like traffic volumes, track length and electrification – while at the same time comparing the efficiency of different companies in an industry.

17. In this section, we set out the methodologies used in the two strands of analysis namely:

- econometric top-down benchmarking of Network Rail routes' total maintenance and renewals costs; and
- econometric top-down benchmarking of Network Rail MDUs' maintenance costs.

Econometric top-down benchmarking of Network Rail routes' total maintenance and renewals costs

18. We used a 5-year panel dataset (covering 2011-12 to 2015-16) to analyse the overall maintenance and renewals costs of Network Rail's routes, using the PR13 definitions of routes. This means that we have ten route datasets, rather than the eight that would reflect Network Rail's current structure[3]. While our analysis would have benefited from using a longer panel dataset, the choice of the period covered was dictated by the availability of data as Network Rail only started collecting data at route level from 2011-12.

19. After cleansing the data and handling any errors and/or outliers, we first examined trends in the data, which helped us to understand differences in routes' spending behaviours and characteristics. We then applied widely used econometric methods of corrected ordinary least squares (COLS) and stochastic frontier analysis (SFA) to estimate the cost function and produce notional efficiency scores for each route. The key independent variables used in our cost function are track length, traffic density (i.e. train km divided by track km), and the average number of tracks (i.e. track km divided by route km).

20. Specifically, we estimated a number of variants of the following model equation:

[3] Our PR13 determination and associated reporting requirements was with respect to ten Routes. Network Rail merged East Midlands and LNE, and separately Kent and Sussex, so that for most of CP5 it has had eight geographical routes. However, as Network Rail's regulatory accounts report on the basis of ten Routes, we have undertaken our analysis on that basis to minimise the risk of errors. In addition, the FNPO was not defined as a Route in PR13 and is excluded from our analysis because it does not have maintenance or renewals costs.

LnCost= f (Ln length of track + Ln traffic density + Ln average number of tracks+ Dummy for final year of CP4 + Time) +error term

21. Where **Ln** means ‘natural logarithm’. Although we based this model on published research and our own previous analysis, data availability also dictated the choice of variables. As earlier mentioned, Network Rail started collecting data at route level in 2011-12 and, unfortunately, data was not consistently collected on all variables that we would wish to include in this analysis (e.g. assets’ age and condition, topography, etc.)

22. Our main model showed that total maintenance and renewal cost (TOTEX) for each of the operating routes *(i)* at time period *(t)* is a function of track length (TRACKKM), traffic density (TRAINTRA), average number of tracks (AVTRACK), a dummy for the final year of CP4 (DYR3), a time trend (T) and a random error:

$$LnTOTEX_{it}=\beta_o+\beta_1 lnTRACKKM_{it}+\beta_2 lnTRAINTRA_{it}+\beta_3 lnAVTRACK_{it}+\beta_4 DYR3+\beta_5 T+e_{it}$$

23. Our preferred model combined maintenance and renewals in total cost, but we also ran and presented models with maintenance and renewals costs separately (see details in Appendices A and B). Our main reason to prefer the combined approach is that it helps to account for the close substitution of some maintenance and renewal activities. We conducted various statistical tests and they all concluded that our model specification is valid. We consider that the model is robust from an econometric perspective. It is also robust to changes in modelling approach (i.e. COLS vs SFA) and small changes to the underlying data.

Econometric top-down benchmarking of MDUs’ maintenance costs

24. Maintenance delivery units (MDUs) are operating units within routes that are responsible for the majority of Network Rail’s maintenance activity. At the time of our analysis, there were 37 MDUs. MDUs accounted for nearly 70% of total network maintenance expenditure during the two years covered by this analysis i.e. 2014-15 and 2015-16.

25. We started our analysis by examining trends in the available two-year data to understand MDUs’ behaviours and characteristics.

26. As for the route analysis, we then estimated the maintenance cost function and produced individual efficiency scores for each MDU using corrected ordinary least squares (COLS) methodology. We mainly based our model’s specification on two strands of previous analysis of MDUs:

- Network Rail's own internal analysis (2010, 2011, and 2012) that used regression analysis with a wide variety of variables to benchmark 39 MDUs. The analysis produced results that, according to Network Rail (2012) "compared the best in class with the other delivery units". Those results were used to set MDUs budgets with efficiency targets for the last three years of CP4.
- The published paper by Wheat and Smith (2008) which applied ordinary least squares (OLS) to a cross section dataset on 53 MDUs in 2005-06 to estimate the marginal cost of running more or less traffic on a fixed network in UK.

27. We conducted a number of statistical model specification tests to check the validity of our model. They all concluded that our model specification is valid. The model is as follows:

    Ln(Maintenance Total Cost)= f(Lntrackkm + Lntraffic density_pax + Lntraffic density_fr + Lnwage + Electrified density+ Speed_ 40-75 density+ Ln average tracks+ Criticality_1 density) +Random error

28. Where:

- ***Ln*** means 'natural logarithm';
- **track km** is the length of the track;
- **traffic density_pax** means passenger train km divided by track km;
- **traffic density_fr** means freight train km divided by track km;
- **average track** stands for track km divided by route km;
- **wage** stands for average real weekly earnings;
- **electrified density** is the proportion of track that is electrified;
- **speed_ 40-75 density** is the proportion of track with speed between 40-75 miles per hour; and
- **Criticality_1 density** is the proportion of track in criticality band 1[4].

29. Given the small size of the two datasets and our inability to obtain data on some theoretically important cost drivers (such as assets' condition / age, topography, etc.), we based our conclusions on our results from the simpler but widely used COLS

[4] Network Rail defines **Route criticality** as a "measure of the consequence of the infrastructure failing to perform its intended function, based on the historic cost of train delay per incident caused by the track asset". Using this measure, each strategic route section (SRS) of the network has been assigned a route criticality band from 1 to 5.The lower the number of the criticality band, the more a delay is likely to cost should infrastructure fail. The classification of each SRS into criticality bands is used in the development of Network Rail's asset policy as a first step to matching the timing and type of asset interventions.

methodology. In our view, COLS is the most appropriate methodology given the size and the quality of the data at our disposal.

3. Findings

30. Table 1 below summarises the modelled *notional* efficiency scores for each route while Table 2 summarises modelled *notional* efficiency scores for each MDU. We describe these as notional efficiency scores because we recognise we are not yet sufficiently confident in the modelling approaches to conclude that the differences represent *bona fide* differences in efficiency as opposed to structural differences between routes, which, because we lack data on them, the models are unable to distinguish from inefficiency.

31. It is also important to understand what numbers in those tables show and how they compare to the efficiency numbers we talk about in our assessment of Network Rail's costs and income, as part of our draft determination. In particular, it is useful to distinguish between two baselines against which efficiency can be measured. First, there is the level of cost that a fully efficient company would incur, given current technology, when delivering the outcomes required for control period 6 (CP6). This is often referred to as 'frontier efficiency', and is a largely theoretical concept. Second, there is the level of cost that we consider Network Rail – given its current performance and current technology – can reasonably be expected to deliver. In the context of a public sector organisation, where it is particularly important to set challenging but ultimately realistic efficiency targets (not least to provide effective reputational incentives), it is the second of these that we are focusing on.

32. Reflecting this, our draft determination focuses on the question of what level of efficiency challenge we think it is reasonable to set Network Rail's management, given where the company is in terms of its ongoing transformation.

33. However, the econometric analysis we have undertaken produces efficiency metrics that compare an estimate of current notional route efficiency to a modelled estimation of how efficient a route could be, given the data available. This is roughly analogous to comparing the efficiency of a route to an estimate of 'frontier efficiency'.

34. In addition, given that our conclusions are based on COLS methodology which adopts a strong (and perhaps unrealistic) assumption that all the deviation from the frontier reflects inefficiency, we made the assumption (as in our PR13 analysis) that 25% of modelled inefficiency is explained by random noise in the data. We therefore applied a 25% uplift to our modelled notional efficiency scores. This practice of uplifting COLS modelled efficiency scores is widely used by other regulators. For instance, in their 2007-2008 relative efficiency assessment, Ofwat adjusted COLS

residuals for water by 10% and for sewage by 20%. For more details about this adjustment, see appendix A and B.

35. Our final modelled notional efficiency scores are presented in Table 1 and Table 2 below.

36. Furthermore, our analysis faced some constraints that should be borne in mind when considering these results. These include the small size of the dataset; unavailability of data for potentially important cost drivers; and (in our route analysis) our inability to fully adjust for the year-on-year fluctuations in renewals expenditure. This means that our modelled cost function may not exactly reflect the true cost structure. Consequently, our view is that we cannot use these results alone to draw strong conclusions about individual route / MDU efficiency levels. However, we consider that they are robust enough to be used to sense-check and support results from other analyses.

37. Our preferred model (COLS) produces a wide range of notional efficiency scores for both routes and MDUs. It estimates that routes and MDUs are on average 84% as notionally efficient as the modelled 'frontier efficient route'. This notionally means that, all other things being equal, an average route/MDU could spend an average of 16% less of its budget if it were as efficient as the 'frontier efficient route'.

Table 1 - Modelled route notional efficiency scores

Year	Anglia	EM	Kent	LNE	LNW	Scotland	Sussex	Wales	Wessex	Western	Average
2011-12	0.83	0.89	0.84	0.85	0.82	0.84	0.98	0.71	1.00	0.90	0.87
2012-13	0.87	0.79	0.81	0.82	0.78	0.86	0.81	0.70	0.96	0.70	0.81
2013-14	0.94	0.91	0.78	0.87	0.77	0.88	0.79	0.81	0.91	0.76	0.84
2014-15	0.80	0.99	0.72	0.91	0.83	1.00	0.80	0.87	0.79	0.76	0.85
2015-16	0.88	0.95	0.74	0.85	0.76	1.00	0.80	0.86	0.86	0.77	0.85
Average	0.86	0.91	0.78	0.86	0.79	0.92	0.84	0.79	0.90	0.78	0.84

Table 2 - Modelled individual MDUs notional efficiency scores

Route	MDU	2014-15	2015-16	Average
Anglia	Ipswich	0.72	0.78	0.75
	Romford	0.78	0.85	0.82
	Tottenham	0.84	0.93	0.89
EM	Bedford	0.81	0.85	0.83
	Derby	0.87	0.92	0.89
LNE	Doncaster	0.97	0.90	0.93
	Leeds	1.00	0.93	0.96
	Newcastle	0.81	0.71	0.76
	Peterborough	0.96	0.96	0.96

Route	MDU	2014-15	2015-16	Average
	Sheffield	0.90	0.89	0.90
	York	0.95	0.87	0.91
LNW	Bletchley	0.78	0.81	0.79
	Euston	0.69	0.78	0.73
	Lancs & Cumbria	0.63	0.63	0.63
	Liverpool	0.92	0.89	0.90
	Manchester	0.72	0.75	0.73
	Saltley	0.90	0.88	0.89
	Sandwell & Dudley	0.82	0.77	0.79
	Stafford	0.99	0.93	0.96
Scotland	Edinburgh	0.93	0.88	0.90
	Glasgow	0.97	0.95	0.96
	Motherwell	0.87	0.87	0.87
	Perth	0.97	0.97	0.97
Sussex	Croydon	0.78	0.79	0.78
	Brighton	0.81	0.80	0.80
Kent	Ashford	0.72	0.78	0.75
	London Bridge	0.81	0.84	0.82
	Orpington	0.89	0.92	0.90
Wales	Cardiff	0.63	0.63	0.63
	Shrewsbury	0.90	0.85	0.87
Wessex	Clapham	0.81	0.83	0.82
	Eastleigh	0.71	0.72	0.72
	Woking	0.78	0.90	0.84
Western	Bristol	0.79	0.81	0.80
	Plymouth	0.81	0.83	0.82
	Reading	0.81	0.79	0.80
	Swindon	0.85	0.78	0.81
	Average	0.83	0.84	0.84

4. Conclusions – Econometric top-down Benchmarking

38. Compared with our international approach in previous periodic reviews, our PR18 econometric top-down benchmarking has been entirely intra-Network Rail. Being able to benchmark Network Rail's routes and MDUs is an important step in achieving our policy objective of moving towards a more route-focused regulation.

39. As this was the first time that we undertook benchmarking at such a disaggregated intra-Network Rail level, we have faced some data quantity and quality constraints. Nevertheless, this analysis forms an important part of our approach to route-level regulation as it gives us a broad understanding of how routes and MDUs have been performing. Moreover, it forms a good basis for future analysis as it identifies technical issues (including data quality) that we need to tackle to ensure our future analysis is more robust and reliable.

40. However, recognising the constraints on this analysis, we have decided that we cannot make firm decisions on Network Rail's CP6 efficiency targets based solely on these results. Our work is strongly indicative that inefficiencies exist within Network Rail but the current constraints on data limit our ability to quantify these with sufficient accuracy. We will use the results to support and sense-check findings from other analyses including bottom-up approaches and more detailed reviews of headwinds and efficiencies. In CP6, we will work with Network Rail to ensure that the data necessary for top-down benchmarking is consistently collected across a wide set of variables.

Appendices

The appendices contain technical papers which describe ORR's work to benchmark Network Rail's infrastructure maintenance and renewals costs. These are:

- Appendix A – Benchmarking of Network Rail's routes; and
- Appendix B – Benchmarking of Maintenance Delivery Units.

Appendix A - Benchmarking of Routes

Econometric top-down benchmarking of Network Rail's routes' costs

Abstract

As part of the 2018 Periodic Review (PR18), ORR is assessing Network Rail's efficiency using various approaches. This will enable ORR to set Network Rail's challenging but achievable efficiency targets for Control Period 6 (CP6). In this paper, we discuss ORR's econometric top-down benchmarking of route-level maintenance and renewals costs.

Using a 5-year panel dataset (2011-12 to 2015-16), we first present the trends in the data to help understand differences in routes' characteristics and spending behaviours. We then use econometric methods of Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA) to estimate the cost function and benchmark routes' notional cost efficiency. Our preferred COLS model suggests that routes with high traffic have a cost advantage; that track length is positively correlated with cost; and that networks with multiple, rather than single track are cheaper to run. Finally, the analysis produces a range of notional cost efficiency scores, relative to the (theoretical) route efficiency frontier.

Our model estimates that routes are on average 84% as efficient as the modelled 'frontier efficient route'. Our analysis was constrained by the size of the dataset; a lack of data for some important cost drivers; and year-on-year fluctuations in renewals expenditure that we could not fully control for. Therefore, although our model is robust from an econometric perspective, our results cannot be used to set routes' efficiency targets for PR18. However, they can be used to sense-check other analyses which inform that decision, and support the general conclusion that there are significant inefficiencies remaining in Network Rail's SBP forecasts for CP6.

1. Introduction

1. This paper discusses the routes data, the methodology and the findings of our PR18 econometric top-down benchmarking of Network Rail's routes' total maintenance and renewals costs for the period 2011-12 to 2015-16.

2. Econometric top-down benchmarking is a common tool used by regulators (e.g. Ofgem, Ofwat, etc.) in price reviews to inform and hence challenge regulated companies' efficiency assumptions.

Use of econometric benchmarking in PR08 and PR13

3. Since the 2008 periodic review (PR08), we have used top-down benchmarking alongside other efficiency benchmarking methods to assess the scope for Network Rail's cost efficiency improvement, and compared this with other evidence to inform our decision regarding the funding that Network Rail required to deliver the outputs for the next control.

4. In our PR08 and PR13 econometric top-down benchmarking, we used an econometric technique known as (cost) frontier analysis to first estimate the cost function and then analyse Network Rail's efficiency. The estimation of the cost function using econometrics has involved looking at how variations in cost drivers (such as network length, train density, average number of tracks, etc.) are correlated with variations in maintenance and renewals costs.

5. Econometric top-down benchmarking provides the basis to establish a comparator's efficiency relative to its peers. This is done by providing an estimate of the extent to which the comparator (in our case routes) is above the minimum cost (i.e. cost of the most efficient firm in the industry) of providing the current level of service.

6. However, it often suffers from a major technical challenge of requiring data of sufficiently high quantity and quality to produce reliable estimates. This also arises from the fact that there is trade-off between the number of cost drivers that can be included in a top-down benchmarking model and the capacity of that model to produce precise estimates. Moreover, while top-down benchmarking may help identify areas for further investigation and challenge, it does not allow for a qualitative understanding of the reasons behind efficiency differences between comparators.

7. In PR08, we placed considerable weight on the results of international econometric top-down benchmarking which used a panel data to compare Network rail's efficiency with that of similar infrastructure managers in Europe. The results of the econometric benchmarking were extensively compared to (and, in general, strongly supported by) the results from other (non-econometric) bottom-up benchmarking. We used these top-down benchmarking results to set Network Rail's efficiency targets.

8. For PR13, we improved our econometric methodology by working on technical issues that Network Rail raised in PR08 (around the quality of data, the usefulness of international rather than single country data, and the stability of the estimated econometric models). Using a carefully cleansed panel data, we considered different estimation techniques (21 in total) including stochastic frontier analysis (SFA). However, in PR13 we decided at an early stage to use bottom-up benchmarking as the basis of our evidence for determining maintenance and renewals efficiency. We

used international benchmarking as a sense check to give us greater confidence in our detailed bottom-up benchmarking analysis conclusions.

Use of econometric benchmarking in PR18

9. From 2011, Network Rail devolved responsibility for day-to-day operation and management of the railway to its ten geographic routes (Anglia, East Midlands, Kent, LNE, LNW, Scotland, Sussex, Wales, Wessex, and Western). These routes were reduced to eight after merger between LNE and East Midland on the one hand, and Kent and Sussex on the other.

10. We view these developments as an opportunity to extend our analysis in PR18 and beyond by conducting an intra-Network Rail benchmarking. One aspect of this intra-Network Rail benchmarking consists of benchmarking routes in terms of their cost efficiency and this is the subject of this paper. Compared to international top-down benchmarking, intra- Network Rail top-down benchmarking has some advantages, in particular in the consistency of data definitions. In addition, Network Rail's routes operate in broadly similar conditions which means they are relatively easy to compare.

11. As this is the first time that we conduct top-down benchmarking at route level, this PR18 analysis serves as a foundation for route level top-down benchmarking analysis in future periodic reviews. While it gives a broad idea on how routes have been performing, it also highlights the issues that we need to tackle in order to improve the reliability of our future analysis.

12. In this paper, we used a mathematical double log (Cobb Douglas) equation to model the cost function (our main dependent variable being total expenditure i.e. maintenance + renewal costs). We applied econometric techniques of Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA) models to a 5-year balanced panel dataset i.e. covering the period from 2011-12 to 2015-16 for Network Rail's ten routes. We conducted the analysis on ten routes rather than eight in order to be consistent with the regulatory accounts publication. However, our future analysis could be undertaken on the eight operating routes. Moreover, while our analysis could have benefited from having a longer panel, we only used a 5-year panel as Network Rail started collecting data at route level from 2011-12.

13. Before conducting the econometric analysis, we ensured that our data was consistent by conducting various checks including identifying potential outliers, missing or inconsistent data, adjustment for year-on-year fluctuations in renewal expenditure, adjusting the cost data for inflation, etc. Any inconsistency was handled with the help of Network Rail.

14. Given the small size of our dataset as well as difficulties in getting reliable data on some potential cost drivers, we preferred to base our final analysis on the results from the simplest but most widely used model specification i.e. the COLS model. We conducted various robustness and model specification checks to ensure reliability of our results.

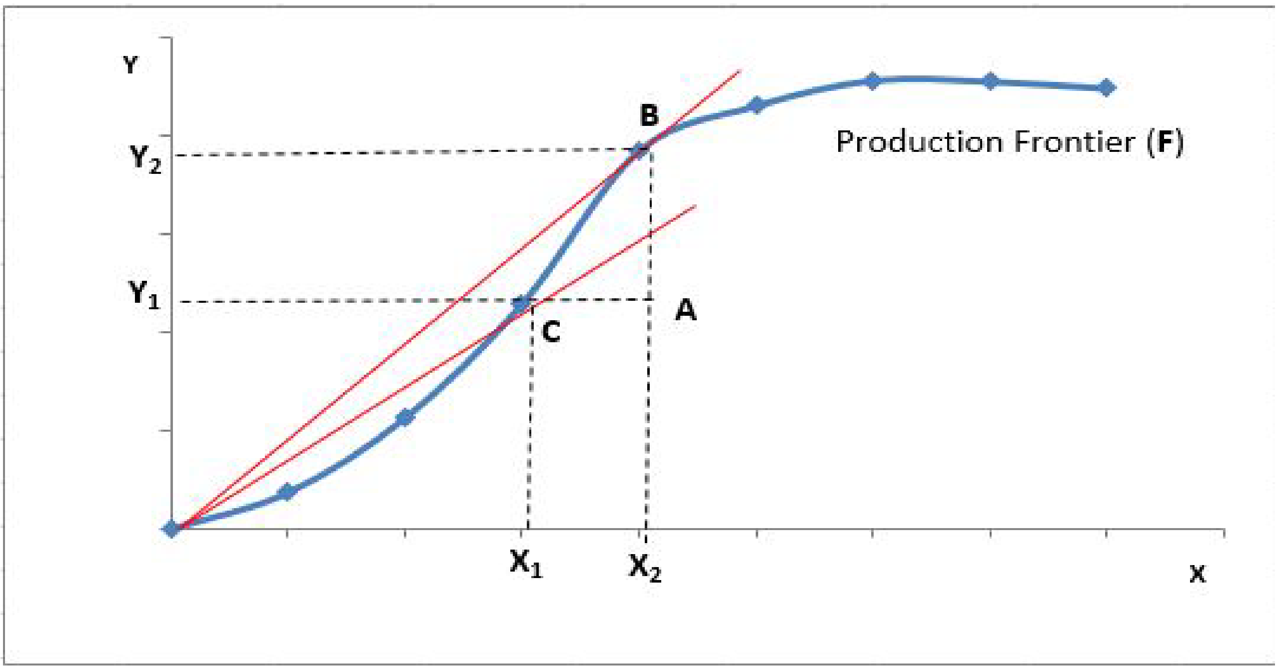
15. This paper is structured as follows: After this introduction, section 2 discusses the general concept of analysing firm's performance, section 3 presents the data used in this analysis as well as different data mining approaches we undertook. Section 4 discusses our cost function, the model specification and the methodology we adopted in this analysis. Section 5 presents our results while section 6 concludes.

2. Firm's performance analysis: productivity, cost frontier and efficiency

16. An increasing number of regulators (including the UK's Ofgem and Ofwat) use top-down benchmarking to inform their understanding of regulated firms' efficiency. Econometric top-down benchmarking can help to determine whether the regulated firm is producing its outputs in the most cost efficient way (i.e. at minimum cost) and to quantify the gap between the actual and the minimum cost.

17. This section introduces the concepts widely used in analysing firms' performance including productivity and efficiency. The following graph represents a firm producing output Y using input X:

Figure 1 Efficiency and productivity



18. In Figure 1 above, line **F** represents the maximum level of production where resources are optimally combined to produce the output under a given production technology. **F** is called the **frontier** of the feasible production. Thus, firms B and C are said to be **technically efficient**. Firm A is inefficient because it could either produce more output using the same amount of inputs (attaining point B) or produce the same quantity of output using fewer inputs (by producing at point C). The distance AB represents the output loss resulting from technical inefficiency. This forms the basis from which the *output–oriented* technical inefficiency is measured. Similarly, the distance AC shows the amount by which the firm can reduce the input but still produce the same level of output. AC forms the basis from which the *input-oriented* technical inefficiency is measured. Therefore, we can understand firm's efficiency as the comparison between its actual production and what it could produce if it produced at the frontier.

19. A firms' **productivity** is the ratio of outputs it produces to inputs it uses. This is a general measure of firm's performance. For a firm producing one output using one input, productivity would be easy to compute. However if the firm produces multiple outputs using several inputs both inputs and outputs must be aggregated in an economically sensible manner in order to produce a **Total Factor Productivity** (TFP) that is a ratio of the two scalars. **Productivity growth** is the difference between growth in output and growth in inputs. Given that production is not always on the frontier, change in productivity can arise from two scenarios: i) movements towards or away from the frontier due to changes in technical efficiency; and ii) shifts in the frontier due to the effect of technological innovations or progress.

20. In Figure 1 above, firm C's productivity (Y_1/X_1) is less than the productivity of firm B (Y_2/X_2). This is because the firm producing at point C is not fully exploiting the economies of scale (as shown by the high slope in the middle of curve F). As the two firms operate at the frontier, they are both considered as technically efficient. However, firm B is said to be more productive as it operates with maximum productivity and maximum efficiency.

21. The economic concept of **duality** implies that we can use the properties of the production function to infer about properties of the cost function. This is because when we say that a firm is maximizing its profit it implies that the firm is therefore minimizing its costs. Thus, we can say that curve F in Figure 1 above also represents the minimum cost at which the maximum output can be produced by combining the inputs given their prices and the existing technology. Therefore, F becomes what we refer to as **cost frontier**. The minimum cost frontier represents the minimum cost relationship between cost (of an activity such as maintenance or renewal in a route)

and the drivers of cost, such as length of track to be maintained (the output), the quality of that output as well as relative prices of inputs used such as labour.

22. For a given level of output, there are two different sources of extra costs that a firm may have to reduce in order to minimize its costs:

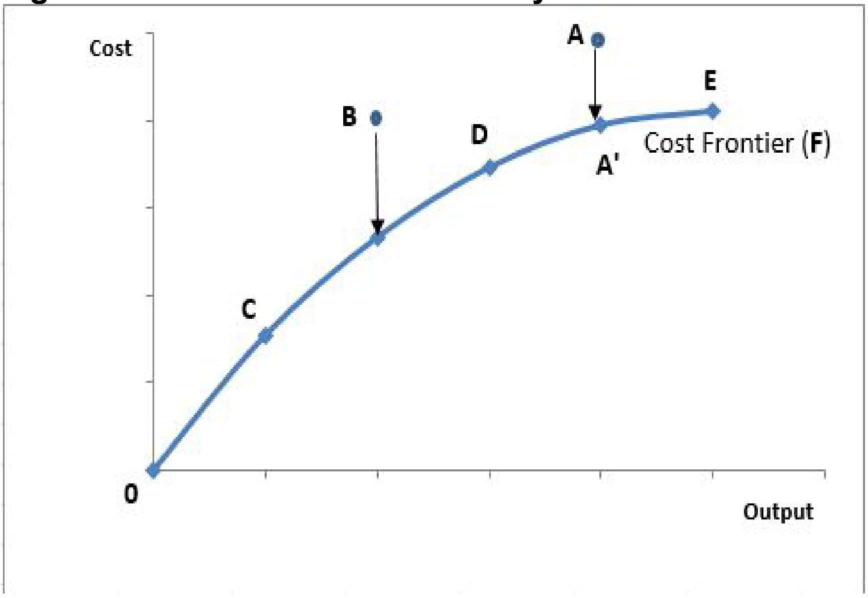
    (i) If there is technical inefficiency in the firm's production process, the firm may work to improve its efficiency by producing the same level of output using fewer inputs;

    (ii) If the same level of output can be produced by a better combination of inputs, then the firm could improve its (allocative) efficiency by choosing the cheapest combination of inputs (this depends not only on the technology but also on the relative price ratios of inputs).

23. **Technical efficiency** shows the proportion of the actual cost which is needed if the firm adopted the best practice (all other things equal). In Figure 2 below, **F** is the cost frontier. Firms C, D and E are efficient as they produce at minimum cost while firms B and A are inefficient as they produce above the frontier. As earlier discussed, by adopting best practice, firms A and B can reduce costs without sacrificing output.

24. Technical efficiency is calculated as the ratio between minimum cost and actual cost. We call that number the **efficiency score** and it helps us to measure the degree to which firms are fully efficient. This is a number between 0 and 1. While 1 indicates that the firm is fully efficient, anything less than 1 indicates that the firm (in our case a route) could continue to maintain and renew a given amount of assets with the same quality but at lower cost. For instance, an efficiency score of 0.7 (or 70%) indicates that a route could potentially reduce its current cost by 30% and still maintain and renew the same amount of assets with the same level of quality.

Figure 2 Cost frontier and efficiency



Cost efficiency (TE) = Minimum cost/Actual cost: TE_A=OA'/OA

0<=TE<=1

3. The route data

25. Before conducting our empirical analysis, it is important to understand the kind of data that is available to us. This section briefly discusses the sources of the data as well as the types of data testing approaches that we adopted to get the final dataset. We then use visual presentation to discuss trends in our data and try to understand their meaning vis à vis our current understanding of routes' operations.

26. As earlier mentioned, we have used a balanced 5-year panel data i.e. from 2011-12 to 2015-16 for the ten routes that existed in that period (Anglia, East Midlands, Kent, LNE, LNW, Scotland, Sussex, Wales, Wessex, and Western). We obtained all the data from Network Rail who compiled it from different sources including Network Rail's regulatory accounts, its asset management services, the annual returns, Network Rail's track access billing team and Network Rail's finance department.

27. We undertook a comprehensive review of the data to ensure its quality and consistency. The three main methods we used to identify outliers and other inconsistencies in the data were:

    (i) The percentage change year-on-year;

    (ii) The visual inspection of trends using charts; and

    (iii) The number of standard deviations from the mean whereby points greater than two standard deviations from the mean were considered outliers.

28. We supplemented these techniques with our understanding of expected behaviours. We used GDP deflator to adjust all the expenditure data for inflation.

29. Whenever we identified an outlier or an inconsistent and/or implausible data point, we shared our concern with Network Rail's colleagues who either provided an explanation for the inconsistency or corrected the error. This allowed us to correct all inconsistencies and challenge all outliers. This gave us a balanced 5-year panel dataset.

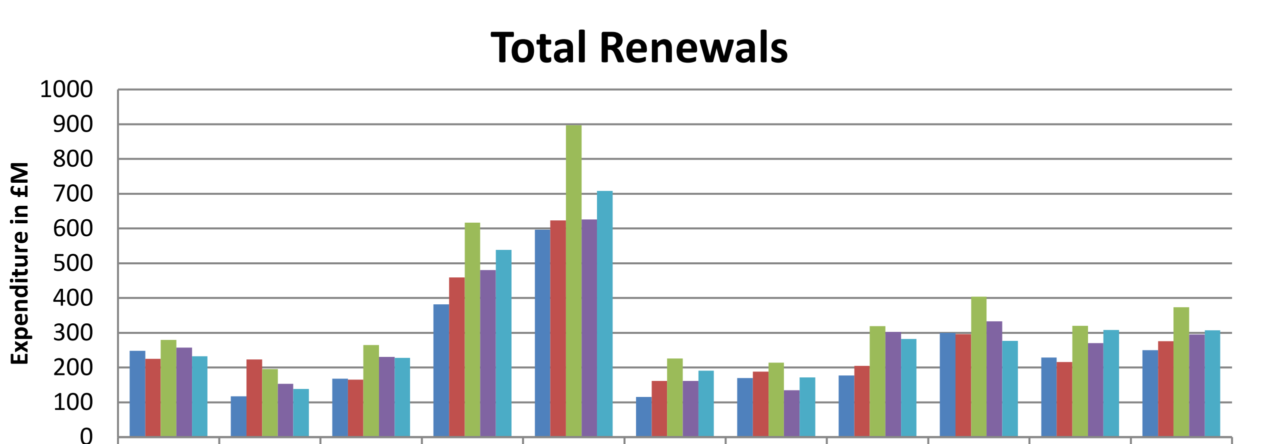
3.1 Visual inspection of trends in the data

30. As this was the first time that we undertake the analysis at route level, it was crucial to start by understanding how different the routes are both in their characteristics and their spending behaviours. This section presents part of the visual inspection that we conducted on the main variables. The years in the charts are presented from 1 to 5 i.e. from 2011-12 to 2015-16 and costs are based on 2015-16 prices. As earlier mentioned, we obtained all the data from Network Rail.

31. Network Rail compiled the data on **Total renewal cost** from its regulatory accounts. Renewals refer to capital expenditure aimed at replacing assets like-for-like following the end of their lifetimes. Renewals expenditure is inherently lumpy and its year-on-year variation is expected to be high[5]. Figure 3 below shows that: LNW and LNE were always the highest spenders on renewals; EM, Kent, Sussex, Wales, Scotland always spent below the average; on average, all routes (except EM) spent more in 2013-14 (end of CP4); annual expenditure varies a lot for each route as expected. For the 5 years coved by our sample, renewals expenditure was 73% of total expenditure on maintenance and renewals. We also observed that on average, the amount spent on track renewal was 29% of total renewals expenditure.

[5] Later in this section, we will discuss how we handled this lumpiness by conducting steady state adjustment.

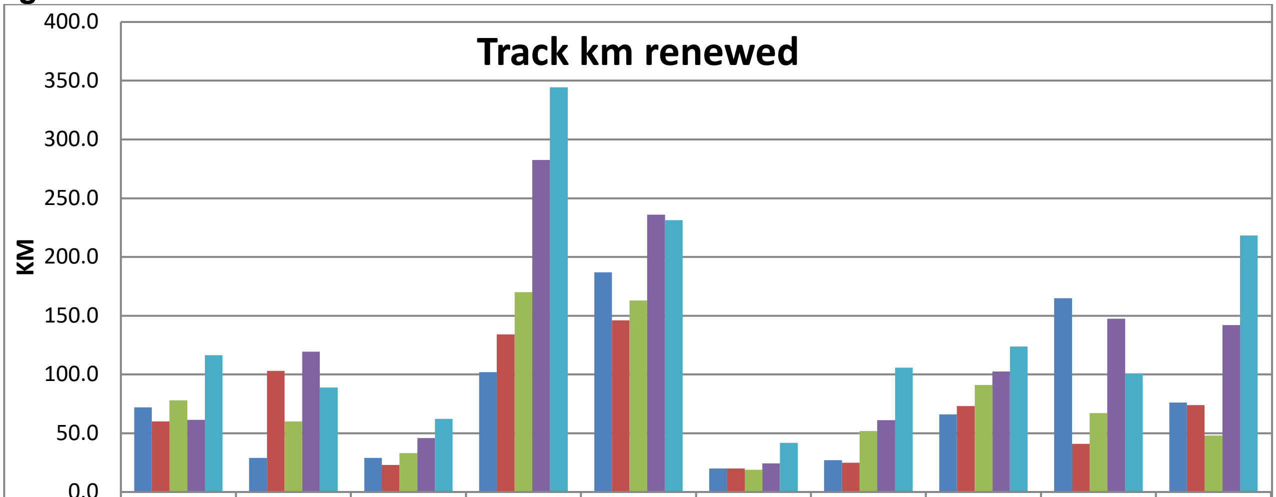
Figure 3 Total renewals cost



	Anglia	EM	Kent	LNE	LNW	Sussex	Wales	Wessex	Western	Scotland	Average
1	247.9	117.1	167.7	381.9	597.1	116	169.8	177.2	299.6	228.9	250.3
2	225.2	223.2	165.3	459.8	623	161.2	188	204.6	296.5	215.9	276.3
3	279.5	196.2	265.3	617	897.5	225.7	214.5	319.2	404.5	320.2	374.0
4	257.2	153.1	231.2	480.4	626.5	162.1	135.1	302.2	333.3	270.2	295.1
5	233	139	228	539	708	191	172	282	277	308	307.7

32. We also examined the data on **track renewal volumes** that Network Rail obtained from its annual returns dataset. We observe that in the last two years, nearly all the routes increased the volumes of track they renewed; LNE and LNW renewed more than other routes while Sussex and Kent renewed the least. As expected, the amount of renewal volumes fluctuates year-on-year.

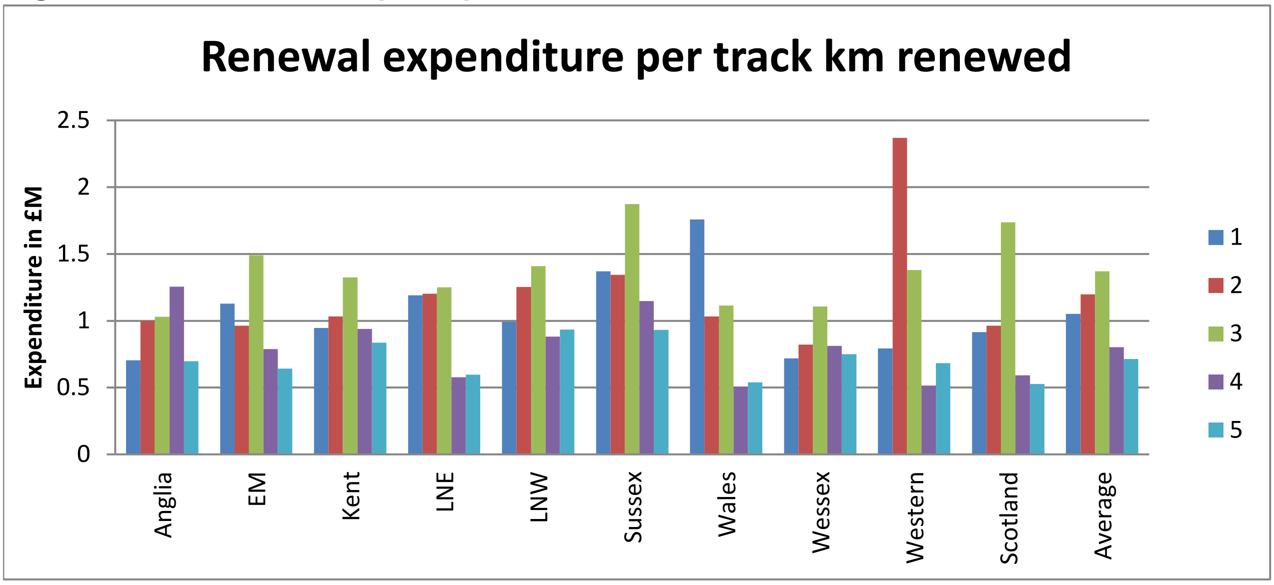
Figure 4 Track renewal volumes



	Anglia	EM	Kent	LNE	LNW	Sussex	Wales	Wessex	Western	Scotland
1	72.0	29.0	29.0	102.0	187.0	20.0	27.0	66.0	165.0	76.0
2	60.0	103.0	23.0	134.0	146.0	20.0	25.0	73.0	41.0	74.0
3	78.0	60.0	33.0	170.0	163.0	19.0	52.0	91.0	67.0	48.0
4	61.4	119.5	45.8	282.5	236.1	24.4	61.0	102.4	147.4	141.9
5	116.3	88.8	62.2	344.3	231.5	41.9	105.8	123.9	101.0	218.3

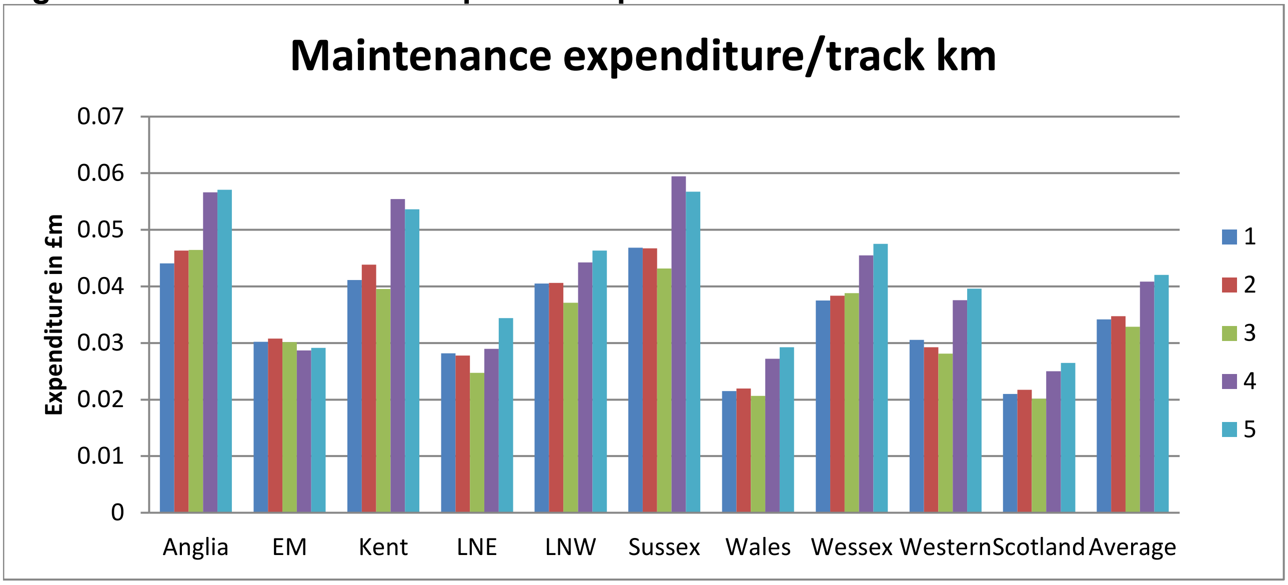
33. Then we examined the **track renewal spend per track km renewed**. Our data reveals that on average in 2013-14 routes spent more money per kilometre of track renewed (about £1.4m/track km) than in any other year. This being the final year of CP4, we suspect this was a reflection of the attempts made by all the routes to meet the CP4 exit targets. On the other hand, for the last 5 years an average route spent about £1.03M per kilometre of track renewed with Sussex always spending above the average. We also observe that Western expenditure increased by 198% between 2011-12 and 2012-13. Network Rail explained this by the fact that Western was the only route using high output equipment during that period. Finally, Wales' expenditure declined by 41% between 2011-12 and 2012-13 and by 54% between 2013-14 and 2014-15.

Figure 5 Track renewal spend per track km renewed



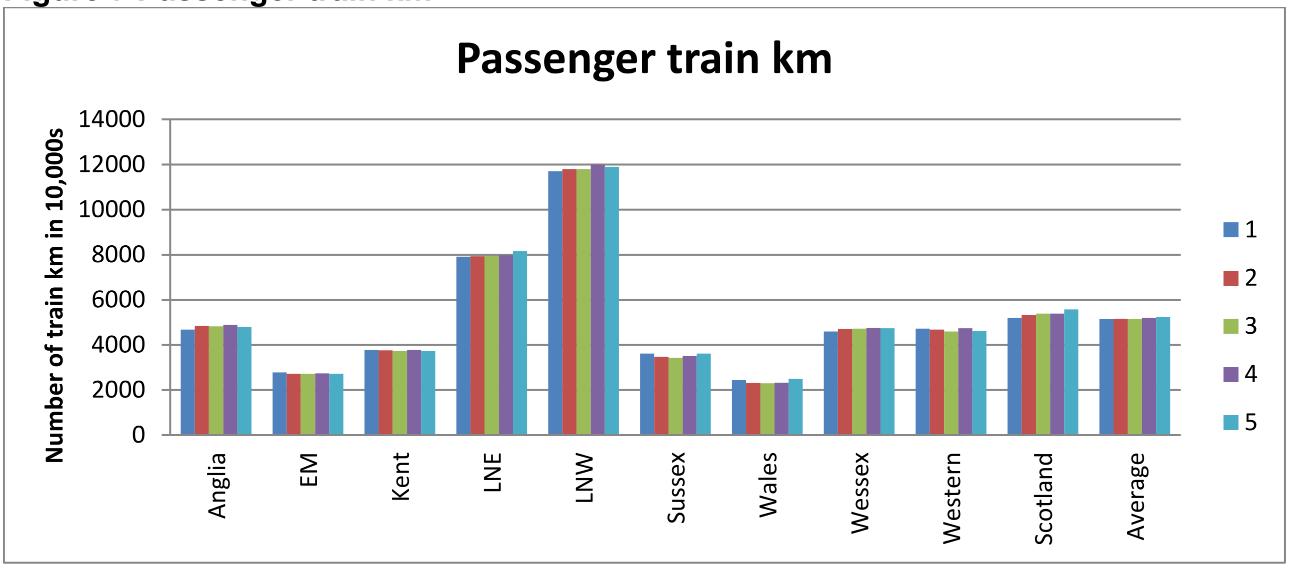
34. Network Rail compiled the data on **maintenance expenditure** from its regulatory accounts. Maintenance activity refers to processes that aim at optimising assets' lifetimes and at sustaining the condition and the capability of an existing infrastructure. We calculated track maintenance expenditure to be £37,000 per year per km of track. We also observed that in the last two years maintenance expenditure per track km increased in almost all the routes. Anglia, Kent, LNW, Sussex and Wessex always spent above the average while during the years covered by this analysis Sussex always spent more than other routes (except Anglia in 2013-14). Scotland, Wales, LNE and EM always spent below the average.

Figure 6 Total maintenance expenditure per track km



35. We then turned to examining traffic density data that Network Rail obtained from its Track Access Billing System. We started by examining **passenger train km**. By comparing data for 2011-12 and 2015-16, we observe that passenger train km has declined in three routes (EM, Kent and Western). It increased by more than 7% in Scotland. On average, passenger train km increased by 1.7% between 2011-12 and 2015-16.

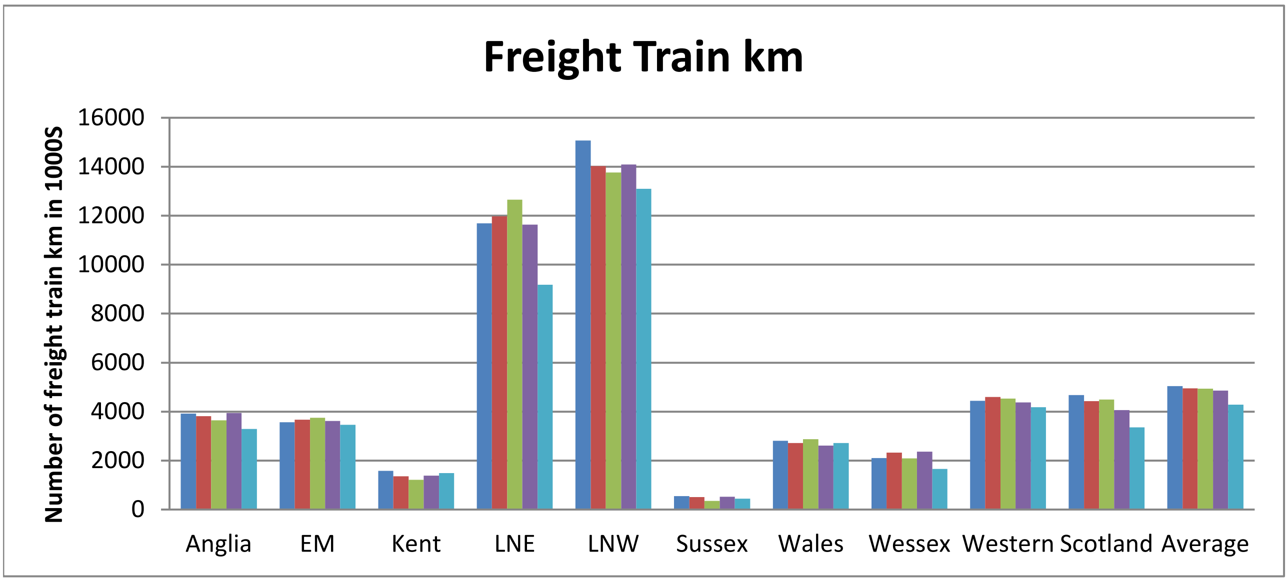
Figure 7 Passenger train km



36. Network Rail also availed to us data on **freight train km** it compiled from its Track Access Billing System. Between 2011-12 and 2015-16, freight train km has declined in all the 10 routes. Comparing freight train km in 2011-12 and 2015-16, it has

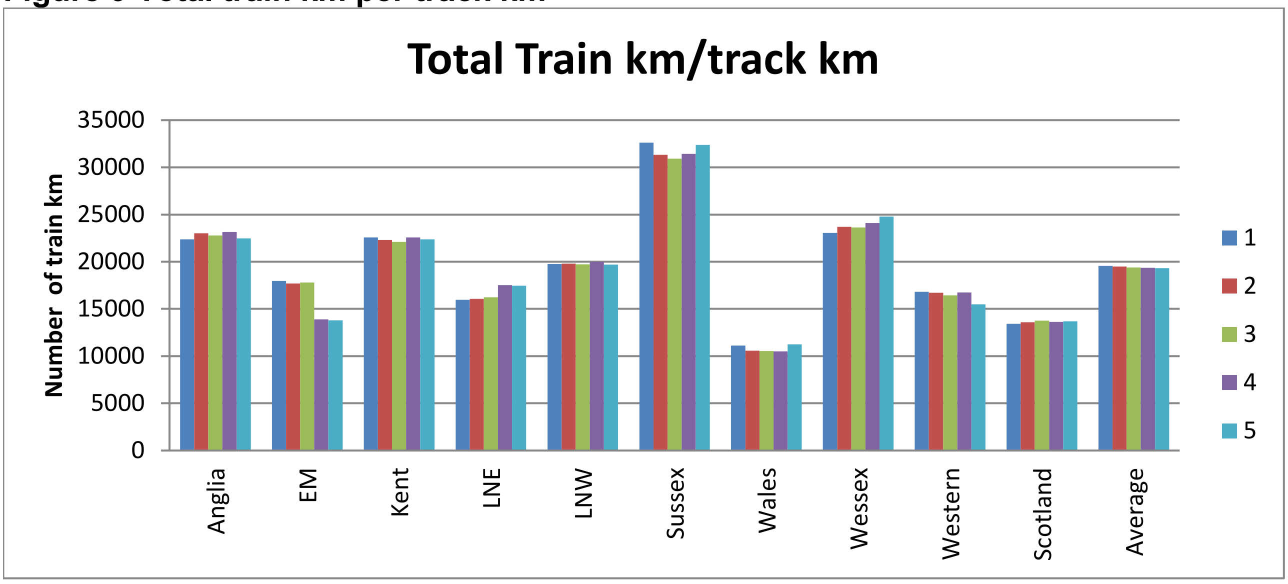
decreased by about 15% on average. This reflects the recent declines in volumes of some commodities transported by rail such as coal.

Figure 8 Freight train km



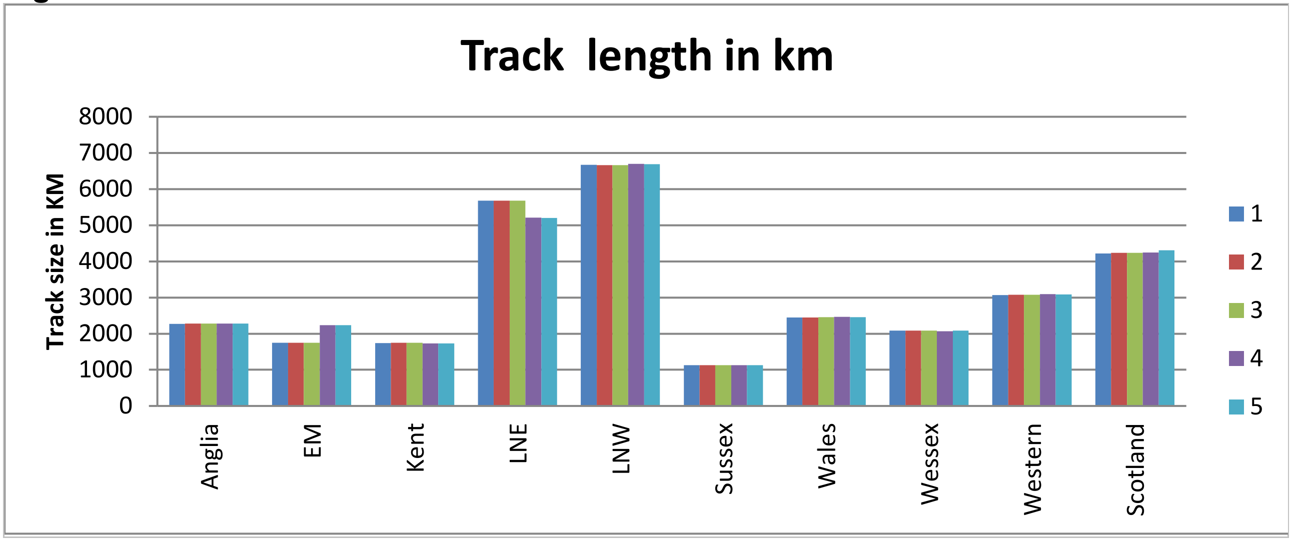
37. We then examined the data on **total train km** (passenger train km +freight train km) **per track km** i.e. **total traffic density**. The data shows that traffic density is almost constant over time in each route, with variations within each route always below 7%. However, levels of traffic density are very different when we compare routes. In the last 5 years, the average traffic density was 19,400 train km per track km. Sussex was by far the most densely used route while Wales and Scotland were the least densely used. Traffic density in the following five routes has been consistently above the average for the last 5 years: Anglia, Kent, Sussex, Wessex and, to a lesser extent, LNW.

Figure 9 Total train km per track km



38. Finally, we looked at the data on **length of track** that Network Rail obtained from its asset management services. As expected, the length of track is almost constant in all routes. In 2014-15, EM's track length increased by 28% but costs and traffic did not follow the same trajectory. We asked Network Rail to explain this. Network Rail argued that this is because the Lincoln area, which was added from LNE, is low cost and low traffic. After our own investigation, we were satisfied by this explanation. The data shows that LNW, LNE, Scotland and Western are the longest tracks while Sussex is the shortest track. Our data shows that EM has the highest number of tracks (i.e. track km per route km) on average (2.5).Average electrification is 47% (zero in Wales, 3% in Western and more than 90% in Kent and Sussex).

Figure 10 Track size



3.2. Dealing with renewals cost fluctuations: steady state adjustment

39. Renewal expenditure is generally lumpy and fluctuates between years for many reasons including those that are not related to changes in cost efficiency.

40. The main goal of steady state adjustment is to adjust a comparator's renewal expenditure for each year to what it would have been had the comparator been renewing at steady state levels. If no adjustment is made, econometric modelling may interpret increased renewal expenditure as a fall in efficiency.

41. Indeed, there are many reasons why the fact that a comparator varies its expenditure over time should not always be interpreted as reflecting changes in efficiency. These could include the following:

(i) Assets are originally installed in lumpy investment projects and the renewal of these assets tends to be concentrated over a time period relating to typical asset lives.

(ii) Backlogs (or, in principle overinvestment) may result from poor policy making but also from situations like funding constraints, weather conditions etc.

(iii) Due to many different asset types and their degradation profiles, there will be natural peaks and troughs in renewal.

(iv) It may be cheaper to renew assets early if other assets nearby are being renewed at the same time (for example to spread access costs and mobilisation costs).

42. Steady state adjustment adjusts expenditure to reflect actual volumes relative to steady state volumes regardless of the reason for the difference.

43. However, there are issues introduced by such an adjustment. In particular, it may have unintended side effects, as its failure to account for reasons for fluctuations means that while it helps in dealing with lumpiness of renewals it may lead to the failure of econometric analysis in picking up the full extent of scope inefficiency. Therefore, in our results we present both steady state adjusted and non-steady state adjusted results. This does not affect our conclusions, as results are comparable.

44. Our PR18 steady state adjustment is similar to the one in PR13[6]. In PR13, we conducted steady state adjustment on track renewal expenditure only. In PR18, we intended to extend the adjustment to other types of assets including signalling, civils, electrification, etc., but we could not do so because of lack of reliable data. This lack of reliable data led us to make the very strong (and perhaps unrealistic) assumption that Network rail was renewing at steady state levels in all other assets.

45. In September 2011, Network Rail submitted its initial industry plan (IIP) to ORR where it committed to a policy target of renewing 2.3% of its track every year until the end of CP4. In our PR13 econometric benchmarking, we used this rate (i.e. 2.3%) as our steady state rate and we wanted to adopt the same approach in PR18. However, in CP5, Network Rail changed this renewal volumes target to 2.1%. With respect to adjusting the expenditure data, for data on CP4 expenditure we used the CP4 renewals target, and for data on CP5 expenditure we used the CP5 renewals target.

[6] For more details and examples, see PR13 Efficiency Benchmarking of Network Rail using LICB

46. In our sensitivity analysis we also tested against only using the 2.3% rate (as in PR13) and the results did not change significantly.

47. We divided total renewal cost into two categories namely track and non-track costs. Then we calculated the:

- **Steady state (SS) volume of track renewal (km)** as *steady state rate of renewal x total track length*.
- We then calculated the **scaling factor** as steady state volume of track renewal (km) / actual volume of track renewal (km).
- A volume scaling factor of greater than 1 means that the route was renewing below steady state while a scaling factor smaller than 1 means that the route was renewing above steady state.
- Then we calculated the **steady state adjusted track renewal cost** i.e. *scaling factor x actual track renewal cost*.
- Finally, the **total steady state adjusted renewals cost** was calculated by summing up *steady state adjusted track renewals cost + actual non-track renewal cost*.

48. By applying the same steady state adjustment rate on all the routes, this approach does not take into consideration the effect of traffic on renewal activities. For instance, one would expect a route with higher traffic density to spend more on renewals and maintenance as more traffic causes higher wear and tear. However, this can also be said about different types of routes in terms of access, topography and other forms of complexity. However, we assume that the rate that is set as a target by Network Rail refers to an average route and this is understood to reflect the behaviour of an average route. To cater for all the problems associated with steady state adjustment, we present results from models with and without the adjustment. The results from both models are broadly comparable and do not affect our conclusions.

4. The cost function, our model specification and methodology

49. This section explains our choice of the functional form and discusses various model specifications we have adopted in this analysis. It briefly explains the difference between our different stochastic frontier analysis models and discusses our preferred corrected ordinary least squares (COLS) specification.

4.1. Functional form

50. In cost function analysis, three main types of functional forms have been used in the literature: linear, double log (i.e. Cobb Douglas) and trans log forms. As in PR08 and PR13, our route data analysis uses a Cobb-Douglas function. We chose this functional form for the following reasons:

    (i) It is simple and fits the data very well.

    (ii) Its estimates can be interpreted as elasticities (i.e. the coefficients represent proportionate changes in cost resulting from proportionate changes in each explanatory variable). These are unit free and can be compared to predictions from economic theory, engineering experience, and estimates from other studies.

    (iii) The linear model is less robust to heteroscedasticity, and the trans log model has too many explanatory variables for it to be estimable using this small dataset.

4.2. Model specification

51. Having selected an appropriate functional form, the next step is to decide the dependent variables (cost measures) and independent variables (cost drivers) to include in the model.

4.2.1 Dependent variable

52. As earlier mentioned, we have used a balanced 5-year panel data i.e. from 2011-12 to 2015-16 for ten routes. Similar to our PR13 approach, our main dependent variable is the total cost (TOTEX) which is the summation of maintenance and renewal costs. Network Rail's financial planning is centred on TOTEX view. Therefore, using TOTEX as our main measure of cost offers the advantage of reflecting the approach that Network Rail takes in setting its targets.

53. Moreover, there may be some trade-offs between maintenance and renewal activities which can also be taken care of by aggregating them in one measure. Indeed, both activities are of different nature with maintenance activity being more cyclical and predictable than renewal activity. Maintenance activity refers to processes that aim at keeping a certain level of infrastructure quality or keep existing assets in working condition. Renewals expenditure on the other hand refers to activities that replace assets like-for-like following the end of their lives. However, to reflect the trend in the literature, we also run models with maintenance and renewals as dependent variables separately.

4.2.2 Explanatory variables

54. In econometrics, there is a trade-off between the number of explanatory variables used in the model, and the model's ability to produce precise estimates. This means that including too many explanatory variables in a model may reduce its ability to distinguish their effects. Therefore, we chose our explanatory variables based on the existing literature as well as on the availability of data. Indeed, we were not able to control for some variables that the literature has usually considered as important drivers of costs (such as route complexity, asset condition, etc.), as the data was not available. Similarly, we did not control for input prices such as labour, material and machinery costs mainly because we did not have data disaggregated to maintenance and renewal level. Although this may not have a significant impact as we can assume that routes faced a comparable level of input prices, this assumption may lead to the model not accounting for some individual routes' inefficiencies such as those in their procurement processes.

55. Although we tested many cost drivers, the following table summarises the explanatory variables that we retained and used in our model. It also presents the intuition behind their relationship with our dependent variable i.e. total cost.

Table 3: List of our independent variables

Variable name	Expected relationship	The intuition behind the relationship
Track km i.e. network size *(Trackkm)*	positive	The longer the track in a route the greater is the volume of track to be renewed and maintained. Greater volumes of work imply greater costs all else being equal.
Traffic density (train km/track km i.e. both passenger and freight train km together, *Traintra*). We also control for passenger train km/track km (*passdens)* and freight train km/ track km *(freidens)* as separate cost drivers	positive	All else equal, we expect an additional train on a fixed network to generate additional wear and tear which calls for additional maintenance and renewal activity and thereby additional costs.
Average number of tracks i.e. track kilometres divided by route kilometres *(Avtrack)*	negative	All else constant, operating routes with more tracks eases traffic flow and, therefore reduces pressures on maintenance and renewals costs.
End of CP4 i.e. year 3 dummy (DYR3)	positive	In our previous visual presentation of the data, we noted that during the final year of CP4 all the routes spend more on maintenance and renewals. We argue that the end of the control period could be a cost driver as routes rush to meet the CP exit targets.
Time trend	positive	This captures all the changes that happen over time including change in input prices, technical change, etc. The effect of input prices (and inflation) is expected to

Variable name	Expected relationship	The intuition behind the relationship
		be positive while the effect of technical progress is expected to be negative, all else constant. Given that we only have a 5-year panel where technical change is not expected to be massive, we expect a positive relationship. We assume a linear trend.

4.2.3. Descriptive statistics

56. Table 4 below presents some statistics to describe our main variables. For the five years under analysis, the data shows that an average route spent £396m on maintenance and renewals (TOTEX) activities. The route that spent the least spent £160m (East Midland in 2014-15) while the one that spent the most spent £1,131m (LNW in 2013-14). The data shows that on average, track size is 3,111km with the smallest route having 1,124 km (Sussex in 2012-13) while the longest had 6,697 km (LNW in 2014-15). Total train density has the mean of 19,426 trains per track km with a minimum of 10,493 train km per track km (Wales in 2014-15) and a maximum of 32,601 train km per track km (Sussex in 2011-12). As expected, passenger traffic density (18,069 train km per track km) is on average, greater than freight traffic density (1,357 train per track km). An average route has two tracks.

Table 4 Summary of variables

Variable	Mean	Std. Dev.	Min	Max
Totex	395.87	218.63	160.09	1131.15
Maintenance	109.10	65.49	48.50	310.00
Renewal	286.77	160.86	95.99	883.84
Trackkm	3111.72	1725.28	1124.00	6697.00
Traintra	19425.60	5816.59	10493.21	32601.19
Passdens	18068.58	6030.88	9365.33	32112.68
Freidens	1357.02	574.82	313.16	2258.18
Avtrack	2.02	0.27	1.57	2.71

57. In econometrics, it is important to check the correlation between variables. When correlation between regressors is high, then there is risk of multicollinearity (i.e. one regressor can be linearly predicted from the others with a substantial degree of

accuracy). In the presence of multicollinearity, although OLS estimates are BLUE[7] it is difficult to conclude on the significance of variables because confidence intervals for coefficients tend to be very wide while t-statistics tend to be very small, so that coefficients will have to be very large in order to be statistically significant. Looking at the correlation between our dependent variables in Table 5 below, we have little reason to worry about multicollinearity. The following table shows the correlation between our main variables.

Table 5 Correlation between variables

Variable	Totex	Maint	Ren	Trackkm	Traintra	Passdens	Freidens	Avtrack
Totex	1							
Maintenance	0.915	1						
Renewal	0.987	0.837	1					
Trackkm	0.907	0.869	0.879	1				
Traintra	-0.119	-0.053	-0.140	-0.395	1			
Passdens	-0.168	-0.106	-0.185	-0.439	0.996	1		
Freidens	0.555	0.570	0.521	0.608	-0.330	-0.414	1	
Avtrack	0.019	0.100	-0.015	-0.117	0.430	0.388	0.278	1

4.3. Methodology

58. In this section, we present the techniques that we used to analyse the routes' cost efficiency. In the literature, two types of analysis have been used to analyse efficiency. These are **parametric and non-parametric methods**. Parametric methods of efficiency analysis utilize econometric techniques. They rely on a specified functional form of production or cost functions. They typically include **regression** and **Stochastic Frontier Analysis (SFA)**. Non-parametric methods use mathematical programming techniques and do not require specification of a cost function. They include **Data Envelopment Analysis** (DEA). DEA is a linear programming-based method whereby the efficiency of a firm is measured by first estimating the minimum cost necessary to secure its output levels when it is compared with similar firms. Then the efficiency of the firm and its scope for efficiency savings are obtained by calculating the ratio of the estimated minimum cost to the observed cost.

[7] Best Linear Unbiased Estimator

59. Similarly, **unit cost analysis** may be used to benchmark a firm's performance by calculating the average cost per unit of output on the basis that the average is also the 'expected' level of any random variable. However, in reality different companies operate in different conditions so that one would expect their average costs to differ even if they were equally efficient (for instance, a large company may have lower unit costs than a smaller company). It is therefore essential to allow for differences in operating conditions when estimating the 'expected' levels of costs for benchmarking purposes. Regression analysis and stochastic frontier analysis help to separate the impact of these conditions on costs and isolate the effect of cost efficiency. They have the advantage of accommodating multiple cost drivers and can test their respective relevance in explaining the variations in costs.

60. As one of the main objectives of our analysis was to estimate the parameters that describe the relationship between the costs and their identified drivers, we chose to apply parametric techniques. These techniques not only allow us to get cost drivers' elasticities and returns to scale but also help us to predict routes' performance relative to 'frontier efficiency'.

61. Therefore, using regression and stochastic frontier analysis, we estimated a number of variants of the following model equation:

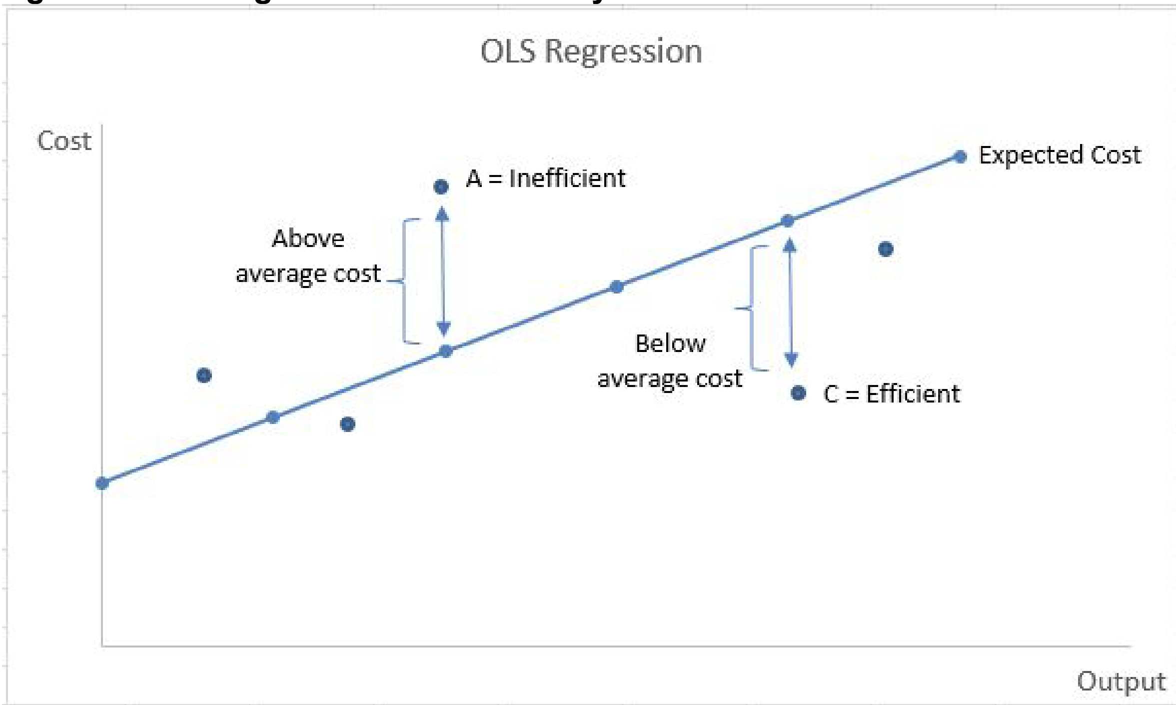
Ln Cost= f (Ln length of track + Ln traffic density + Ln average number of tracks+ Dummy for final year of CP4 + Time) +error term

62. Specifically, our main model is such that the total maintenance and renewals costs (TOTEX) for each operating route *(i)* at time period *(t)* is a function of track kilometre (TRACKKM), traffic density (TRAINTRA), average number of tracks (AVTRACK), dummy for year 3 (i.e. 2013-14) which is the final year of CP4 (DYR3), a time trend *(t)* and a random error:

$$LnTOTEX_{it}=\beta_o+\beta_1 lnTRACKKM_{it}+\beta_2 lnTRAINTRA_{it}+\beta_3 lnAVTRACK_{it}+\beta_4 DYR3+\beta_5 T+ e_{it} \quad (1)$$

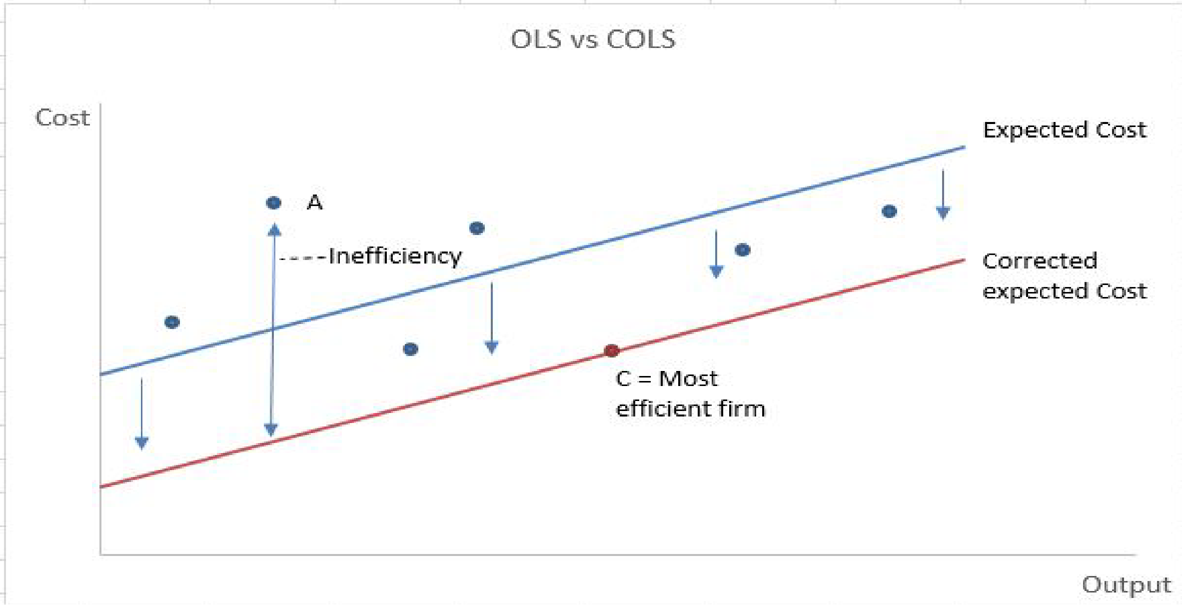
63. When estimating efficiency, it is essential to appropriately specify the error term as it is the error term that yields the information about firm's inefficiency. Under OLS, we estimate a line that passes through the centre of the observed data points i.e. given the information available, the OLS line defines the costs that one would expect an average company to incur given its output. The distance between the OLS line and observed points is the residual (see Figure 11 below).

Figure 11 OLS regression and efficiency



64. Some companies have higher costs (i.e. negative residuals for dots above the line) and some have lower costs (i.e. positive residuals for dots below the line). Therefore, for the purpose of defining “efficient cost”, the OLS line cannot define the ‘**cost frontier**’ as some companies can achieve lower costs. The Corrected OLS (COLS) approach adjusts the model by adding the largest negative OLS residual to the estimate of the intercept parameter. This eliminates the positive residuals as the line moves to the point(s) where the residual equals to zero which constitutes the frontier. Now the distance from the frontier captures notional inefficiency for each firm. See Figure 12 for illustration.

Figure 12 COLS frontier and efficiency



65. COLS has the advantage of being simple and it is widely used by regulators. Given the small size and the quality of our data, we chose to use COLS as our main model. Under OLS (COLS) the consistency of the estimates of elements of β (other than the intercept) does not depend on the assumed distribution of the error term. Moreover, the estimated intercept provides a consistent estimate of the actual intercept. However, COLS' main drawback is that it assumes that any unexplained variation in cost is due to relative (in)efficiency. This assumption is not realistic as many other factors including errors in data measurement, omitted explanatory variables, modelling errors, and other unobservable factors may explain some of the variation.

66. We cater for this draw back in two ways:

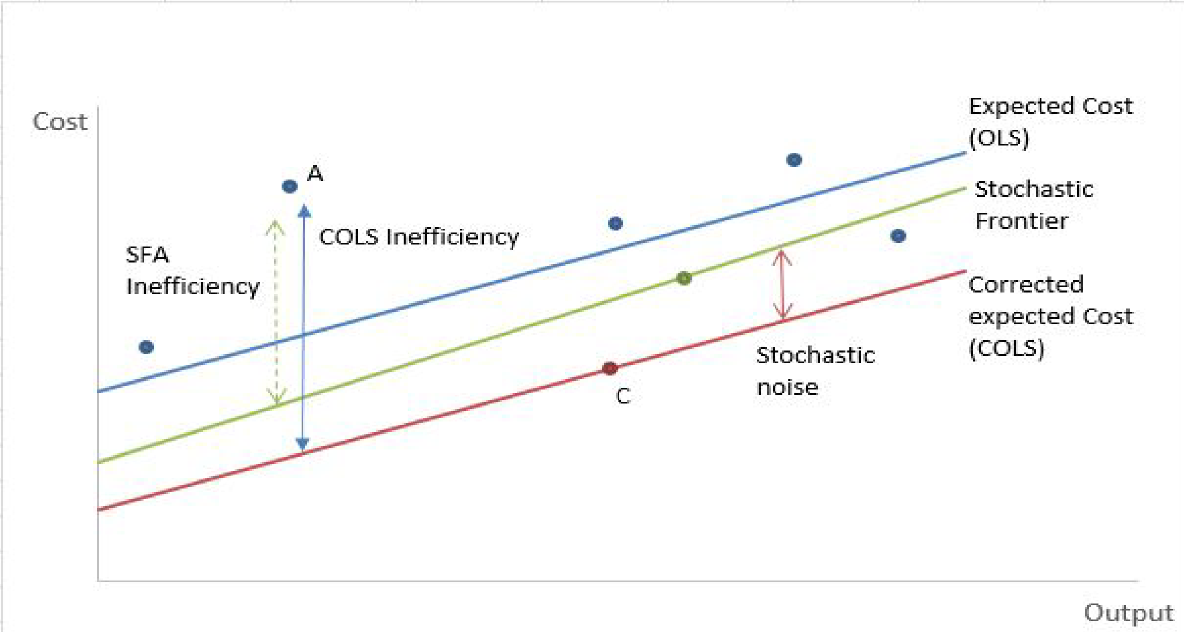
(a) As in PR13, we will assume that 25% of inefficiency is explained by the random noise by applying an uplift of 25% to our efficiency scores. Other regulators have previously used similar methods of adjusting efficiency scores to account for noise. Our approach assumes the existence of some noise in every observation with larger efficiency gaps assumed to have a larger amount of noise. We use the following formula to adjust the scores: $X^*_{i,t}=X_{i,t}+0.25(1-X_{i,t})$, where $X_{i,t}$ is route *i*'s COLS efficiency score in period *t*, and $X^*_{i,t}$ is its "noise adjusted" efficiency score for the same period.

(b) We also applied the stochastic frontier analysis (SFA) technique. The panel stochastic frontier model used to predict technical efficiency has similar specification as OLS except that the error term is decomposed into two independent elements: $v_{it} \sim iid\ N(0, \sigma v^2)$ is the random noise error component and $u_{it} \geq 0$ is the technical inefficiency error component as in equation (2) below:

$$LnTOTEX_{it}=\beta_0+\beta_1 lnTRACKKM_{it}+\beta_2 lnTRAINTRA_{it}+\beta_3 lnAVTRACK_{it}+\beta_4 DYR3+\beta_5 T+v_{it}+u_{it} \quad (2)$$

67. The SFA line usually lies between OLS and COLS and defines an efficiency frontier that allows for particular pattern of stochastic data error. To illustrate this, Figure 13 below compares OLS, COLS and SFA frontiers. In our analysis, we explored different types of SFA, which differ based on their assumption about the error term's distribution.

Figure 13 OLS, COLS and SFA frontiers and efficiency



68. The following table summarises the main differences between our different models based on their assumptions about the error term's distribution.

Table 6 A comparison between different models and their assumptions

Model name	Model's functional[8] form	Components of unexplained variable	Components of inefficiency
OLS/COLS	$LnC_{it}=\alpha+f(X_{it};\beta)+u_{it}$	-Assumes no noise. -Entire residual is interpreted as inefficiency	Inefficiency varies independently across routes and across time
PSFA (Pooled SFA)	$LnC_{it}=\alpha+f(X_{it};\beta)+u_{it}+v_{it}$ $u_{it}\sim\lvert N(0,\sigma_u^2)\rvert$, $v_{it}\sim N(0,\sigma_v^2)$	noise and inefficiency	Inefficiency varies independently across routes and across time

[8] For a detailed discussion about these models, please refer to our previous publication "PR13 Efficiency Benchmarking of Network Rail using LICB"

Model name	Model's functional[8] form	Components of unexplained variable	Components of inefficiency
Pitt and Lee (1981):PL	$LnC_{it} = \alpha + f(X_{it};\beta) + u_i + v_{it}$ $u_i \sim |N(0, \sigma_u^2)|$, $v_{it} \sim N(0, \sigma_v^2)$	noise and inefficiency	Time invariant but perfect correlation over time i.e. inefficiency varies between firms but is constant over time
CUESTA Linear	$LnC_{it} = \alpha + f(X_{it};\beta) + u_{it} + v_{it}$ $u_i \sim |N(0, \sigma_u^2)|$, $v_{it} \sim N(0, \sigma_v^2)$ $u_{it} = exp(\delta_{1i}t).u_i$	Noise and inefficiency	Time varying in a linear trend
CSSRE	$LnC_{it} = \alpha_{1i} + \alpha_{2i}t + \alpha_{1i}t^2 + f(X_{it};\beta) + v_{it} = Z'\alpha i +$ $f(X_{it};\beta) + v_{it}$ $E[\alpha_i] = \alpha,\ V[\alpha_i] = \Omega$	Noise and inefficiency	Time varying with quadratic trends. Treats firm effects as random variables which can be correlated. Assumes no correlation between firm effects and regressors.
RE (Random Effects)	$LnC_{it} = \alpha + f(X_{it};\beta) + u_i + v_{it}$	Noise and inefficiency (random effects are interpreted as inefficiency) but no distribution assumption is made about inefficiency	Time invariant
True Random Effect	$LnC_{it} = \alpha_i + f(X_{it};\beta) + u_{it} + v_{it}$ $u_{it} \sim |N(0, \sigma_u^2|$, $v_{it} \sim N(0, \sigma_v^2)$ $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$	Noise, inefficiency and firm effects interpreted as unobserved heterogeneity	Inefficiency varies independently across routes and across time

5. Our results

69. This section presents and analyses our results. We start by presenting findings from our preferred COLS model using steady state adjusted data. To check the robustness of our findings, we ran various specifications of our preferred COLS model. We also checked our model specification by conducting various tests[9] including the testing for skewness, heteroscedasticity, omitted variables, and multicollinearity. None of the tests concluded that our model specification is invalid.

70. In the second part of this section, we discuss our results from the stochastic frontier analysis (SFA) and compare them with our COLS results. In this section, we also discuss efficiency scores from our COLS model and compare them with those from

[9] Test statistics are not presented here but are available on request

SFA. Finally, we conduct further robustness checks by presenting the results from our analysis of the data that is not steady state adjusted.

71. These results should be considered with caution given the fact that our data as well as our model have some shortcomings that may affect their robustness. These include:

    (i) As earlier discussed, renewals expenditure may fluctuate without this being related to efficiency. Although our objective was to undertake steady state adjustment for all the asset types, we were only able to do this for track renewals expenditure, as the data for other assets was inconsistent or missing. This means that our cost frontier may not exactly reflect the true cost structure. This is expected to improve in future analysis as longer time series may help to smooth out these fluctuations.

    (ii) Our small dataset coupled with lack of data on some important cost drivers such as topography, age of the network/asset conditions, etc. meant that we could not control for them. Similarly, heterogeneity and other factors (such as weather condition) that vary by year are beyond route's control but affect their costs. OLS does not consider this temporal information and does not exploit the benefits of having a panel dataset. This may lead to our preferred model producing too low efficiency scores i.e. over-statement of inefficiency. However, we note that all the model specification tests we conducted showed that our model is properly specified. Moreover, given that there is always a trade-off between controlling for many variables and getting reliable results from a regression on the one hand and the fact that we have a small dataset on the other, it may not have been appropriate to include all the missing cost drivers in this analysis even if we had had the data. In addition, we conducted robustness checks by running models that take into consideration the benefits of a panel dataset. Our results are comparable.

72. One of the main objectives of this analysis was to set a foundation for our future analysis in our efforts to regulate Network Rail at route level. We anticipate that our future analysis will give us more robust estimates as we get longer data series which will enable us to do all the necessary adjustments. As for now, we regard these results as simply indicative of where more efforts could potentially be focused.

5.1 Corrected Ordinary Least Squares (COLS) Model

5.1.1. Our main COLS Model

73. As discussed above, COLS is a basic model that it is commonly used by regulators. Its main advantage lies in its simplicity. The technique consists of estimating the OLS line that passes through the centre of the observed data points and then adjusting

the model by keeping the same gradient as the OLS line but changing the intercept until no firm has observed costs below the line. Table 7 below shows our results from OLS and a summary of COLS efficiency scores (both adjusted with a 25% uplift and non-adjusted). The model has an R-squared value of 0.92. The results suggest that:

(i) Increasing track size and traffic density by 1% leads to an increase in total (i.e. maintenance and renewals) costs by 0.95% and 0.8% respectively;

(ii) There are economies of densities: increasing traffic density increases the cost less than proportionally i.e. most densely used routes have cost advantage. However although our coefficients for track km are consistently below 1 which would suggest the existence of economies of scale, we tested the hypothesis whether the coefficient for track km is statistically different from 1 and found that it is not. Therefore, we did not find evidence to support the existence of economies of scale. Our results points more to the existence of constant returns to scale.

(iii) For a given length of track km, it is cheaper to run it in multiple than single tracks. Increasing the average number of tracks by 1% leads to a cost reduction of 0.34%.

(iv) Year 2013-14 (which is the final year of CP4) is a statistically significant determinant of costs. One explanation for this could be that routes' expenditure decisions were influenced by the fact that they were in the final year of the control period. Our observation on trends in data also confirmed this.

(v) The summary of efficiency scores reports an average difference between the modelled frontier efficiency and route's notional efficiency scores of 21%. However, as earlier discussed, OLS unrealistically considers that all the deviation from the frontier is explained by inefficiency. In our analysis, we assumed that 25% of the score could be explained by noise in the data. We therefore adjusted this score by a 25% uplift as explained above.

(vi) Our model's final score shows that there is an average difference of 16% between the modelled frontier efficiency and routes' notional efficiency scores. This notionally means that, all other things being equal, an average route could have spent an average of 16% less (if it operated at the level of the modelled frontier) and still renewed and maintained the same length of track with the same traffic density and same average number of tracks. However, we are not sufficiently confident in the analysis or the data that underpins it to conclude that these notional efficiency score precisely represent *bona fide* differences in efficiency between routes.

(vii) Table 88 below presents our modelled notional efficiency scores as compared to the modelled 'frontier efficiency'. Although the model behaves technically well, we observe that the inferred routes' efficiency scores fluctuate substantially through the years. This may be a direct consequence of fluctuations in renewals expenditure (and is why modelling of 'Totex' in rail and other industries has proved particularly difficult).

Table 7 OLS results and COLS efficiency scores for our main model

Lntotex	Coef.	Std. Err.	t	P>t	95% Conf. Interval	
Lntrackkm	0.953	0.044	21.9	0.000	0.865	1.040
Lntraintra	0.792	0.087	9.09	0.000	0.616	0.968
Lnavtrack	-0.341	0.178	-1.92	0.062	-0.70	0.018
DYR3	0.214	0.050	4.28	0.000	0.113	0.315
Time	0.004	0.014	0.26	0.796	-0.025	0.032
_cons	-0.008	0.049	-0.16	0.871	-0.106	0.090
Summary of Non adjusted COLS efficiency scores						
	Observations	Mean	Std. Dev.	Min	Max	
eff_cols	50	0.79	0.11	0.60	1	
Summary of adjusted COLS efficiency scores						
	Observations	Mean	Std. Dev.	Min	Max	
Eff_cols_adj	50	0.84	0.33	0.70	1	

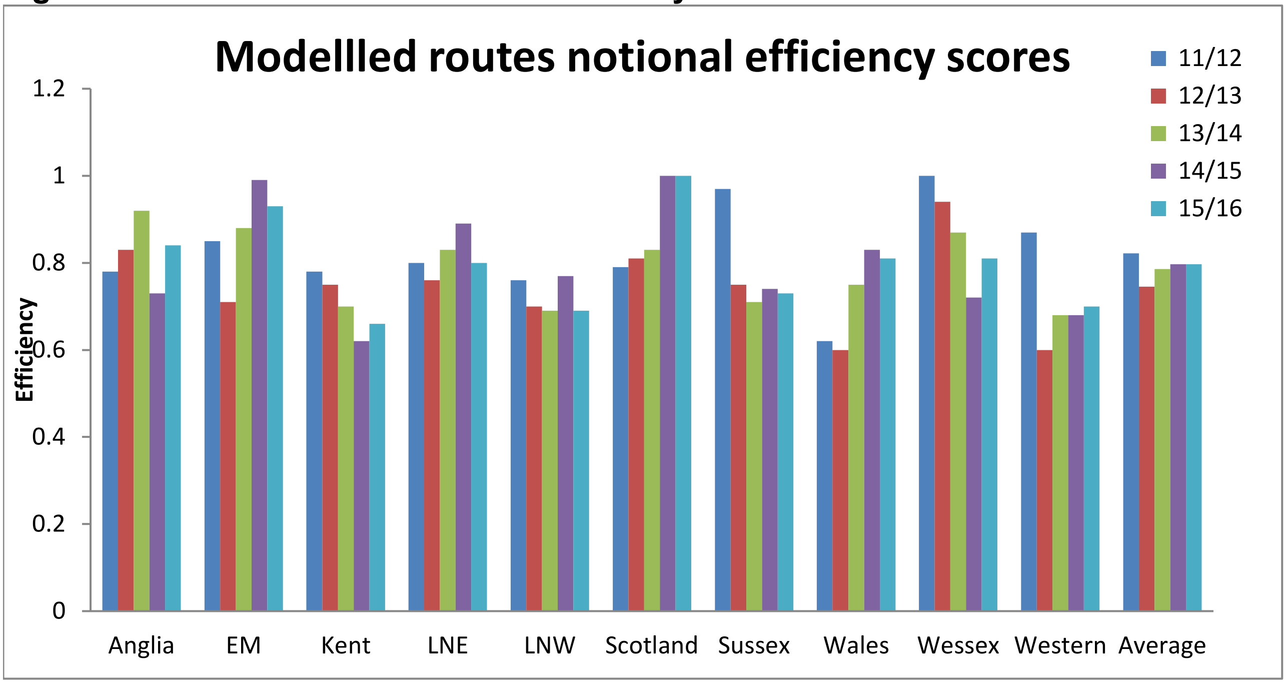
74. In Table 8 below, we show modelled notional efficiency scores for each route from our COLS model (these results include the 25% noise adjustment).

Table 8 Modelled notional efficiency scores

Year	Anglia	EM	Kent	LNE	LNW	Scotland	Sussex	Wales	Wessex	Western	Average
11/12	0.83	0.89	0.84	0.85	0.82	0.84	0.98	0.71	1.00	0.90	0.87
12/13	0.87	0.79	0.81	0.82	0.78	0.86	0.81	0.70	0.96	0.70	0.81
13/14	0.94	0.91	0.78	0.87	0.77	0.88	0.79	0.81	0.91	0.76	0.84
14/15	0.80	0.99	0.72	0.91	0.83	1.00	0.80	0.87	0.79	0.76	0.85
15/16	0.88	0.95	0.74	0.85	0.76	1.00	0.80	0.86	0.86	0.77	0.85
Average	0.86	0.91	0.78	0.86	0.79	0.92	0.84	0.79	0.90	0.78	0.84

75. This information can also be visualised in Figure 14 below.

Figure 14 Modelled routes notional efficiency scores



5.1.2. Other COLS models

76. We run different COLS model specifications. Table 9 below compares their results with our main model (TOTEXOLS) we discussed above. In model MAINOLS, our dependent variable is maintenance cost while model RENOLS uses renewals cost as dependent variable. The following three models (TOTPAFR, MAINTPAFR and RENPAFR) are similar to the previous ones but use passenger and freight traffic densities as separate regressors instead of a combined traffic density variable (TRAINTRA).

Table 9 Various COLS model specifications

Variable	TOTEXOLS	MAINTOLS	RENOLS	TOTPAFR	MAINTPAFR	RENPAFR
Lntrackkm	0.953***	0.999***	0.939***	0.984***	0.951***	0.999***
Lntraintra	0.792***	0.862***	0.783***			
Lnavtrack	-0.341*	0.063	-0.492**	-0.209	-0.039	-0.276
Time	0.004	0.060***	-0.019	0.001	0.063***	-0.024
DYR3	0.214***	-0.140***	0.333***	0.216***	-0.136***	0.335***
Lnpassdens				0.716***	0.855***	0.681***
Lnfreidens				-0.015	0.133**	-0.067
_cons	-0.008	-0.107**	0.026	0.007	-0.102**	0.045
N	50	50	50	50	50	50
r2	0.92	0.94	0.88	0.92	0.95	0.88
Eff. Score	0.84	0.84	0.80	0.83	0.84	0.78
*** p<.1; ** p<.05; *** p<.01**						

77. The results in the above table suggest that there are economies of density. They also show that passenger traffic density is a statistically significant determinant of maintenance and renewals costs.

78. However, freight traffic density is only positive and significant in the model with maintenance as the dependent variable. In other models, it is negative, very small and not significant which seems counter intuitive. Freight trains are usually slower and heavier than passenger trains. Therefore, the two types of trains have slightly different wear and tear characteristics on the network and thereby different cost implications. However, we cannot fully isolate the effect of each as opposed to the other as in each there is an element of speed and weight affecting wear and tear of the network they both run on. Moreover, in our data, freight traffic is only 8% of passenger traffic. This means that given that both are intended to measure the effect of traffic density on the costs of wear and tear, it may be that the passenger traffic variable dominates the freight traffic variable when we control for both together. Furthermore, given that freight is a small fraction of train km, this may imply that for the same marginal cost (given average cost is greater by definition), the elasticity for freight should be smaller than the one for passenger traffic. We therefore consider that aggregating passenger and freight traffics together in one variable (TRAINTRA)

is more pragmatic and sensible. Network Rail also recommended this in their comments on our PR13 top-down benchmarking results and during our discussions in the course of this analysis.

79. For a robustness check, we also ran the same models as above but using the data where renewal expenditure is not steady state adjusted. We obtained comparable results as presented in Table 10 below.

Table 10 COLS models with renewals cost data that is not steady state adjusted

Variable	TOTEXOLS	MAINTOLS	RENOLS	TOTPAFR	MAINTPAFR	RENPAFR
Lntrackkm	0.964***	0.999***	0.954***	0.936***	0.951***	0.934***
Lntraintra	0.761***	0.862***	0.728***			
Lnavtrack	0.127	0.063	0.144	0.088	-0.039	0.129
Time	0.050***	0.060***	0.046***	0.051***	0.063***	0.047***
DYR3	0.183***	-0.140***	0.287***	0.186***	-0.136***	0.290***
Lnpassdens				0.736***	0.855***	0.696***
Lnfreidens				0.089	0.133**	0.072
_cons	-0.139***	-0.107**	-0.15***	-0.13***	-0.102**	-0.149**
N	50	50	50	50	50	50
r2	0.936	0.941	0.906	0.937	0.948	0.906
Eff. Score	0.88	0.88	0.86	0.88	0.88	0.85
*** p<.1; ** p<.05; *** p<.01**						

80. The main model (TOTEXOLS) shows that track size and traffic density are positive and significant determinants of maintenance and renewal costs at 1% with coefficients 0.96 and 0.76 respectively (comparable to the ones obtained with the steady state adjusted data i.e. 0.95 and 0.79 respectively). All the model show comparable results. Passenger traffic density is always positive and significant. Freight density has the expected positive sign but its impact remains very small. The main model (TOTEXOLS) shows that on average, and all other things equal, the difference between modelled frontier efficiency and actual notional efficiency is 12% compared to 16% obtained using data with track only steady state adjustment. In these models, average track becomes positive and is not a statistically significant determinant of costs. This may suggest that the model with steady state adjustment is superior to the one with data that is not steady state adjusted.

5.2. Stochastic frontier analysis (SFA)

81. As earlier discussed, the main problem with COLS models is that they assume that any unexplained variation in cost is due to inefficiency. In this section, we apply the SFA models that have the advantage of decomposing the unexplained variation in its two main components i.e. the error term and the inefficiency term. In these models, inefficiency is unobserved, is heterogeneous between comparators, and probably has an element of time persistence or time invariance. As earlier discussed, these models make different assumptions about the inefficiency term and some of them are able to make a distinction between efficiency and other unobserved time invariant factors.

82. In all the six SFA models as shown in Table 11 below, track size and traffic density are positive and statistically significant determinants of maintenance and renewal costs. All the results are comparable to our COLS results. Apart from CUESTAL (for traffic density), all the models confirm the existence of economies of densities for traffic. Average number of tracks is statistically significant in three SFA models but has always the expected negative sign. Similarly, the dummy for year 2013-14 is positive and statistically significant in five out of six SFA models. In the CSSRE model, it shows a negative but very small coefficient. As expected, modelled notional efficiency scores vary for each model with CUESTAL and CSSRE producing scores that are comparable to our adjusted COLS scores.

Table 11 SFA Models with renewal cost data that is steady state adjusted

Variable	COLS	PSFA	PL	CUESTAL	RE	TRRE	CSSRE
Lntrackkm	0.953***	0.953***	0.963***	0.974***	0.948***	0.948***	0.952***
Lntraintra	0.792***	0.792***	0.863***	1.080***	0.838***	0.817***	0.767***
Lnavtrack	-0.341*	-0.341**	-0.345*	-0.238	-0.412	-0.33	-0.373***
Time	0.004	0.004	0.004	0.023	0.004	0.004	0.223
DYR3	0.214***	0.214***	0.214***	0.225***	0.214***	0.214***	-0.007
_cons	-0.008	-0.008	-0.111**	-0.239**	-0.008	-0.012	0.014
N	50	50	50	50	50	50	50
Eff. Score	0.84	0.99	0.9	0.84	0.91	0.99	0.82
*** p<.1; ** p<.05; *** p<.01**							

83. For a robustness check, we also run SFA models for the data with no steady state adjustment. Again, we get comparable results. Track size and traffic density are

positive and significant determinants of maintenance and renewals costs. The existence of economies of density is confirmed in all models. Contrary to the results from the steady state adjusted data analysis, here the average number of track is not a significant determinant of maintenance and renewal costs while time becomes positive and significant. Again, the dummy for year 2013-14 (final year of CP4) is positive and statistically significant.

Table 12 SFA Models with renewals cost data that is not steady state adjusted

Variable	COLS	PSFA	PL	CUESTAL	RE	TRRE	CSSRE
Lntrackkm	0.964***	0.964***	1.003***	1.030***	0.961***	0.989***	0.952***
Lntraintra	0.761***	0.761***	0.896***	0.942***	0.836***	0.825***	0.828***
Lnavtrack	0.127	0.127	0.226	0.251	0.064	0.122	0.127
Time	0.050***	0.050***	0.050***	0.054***	0.050***	0.056***	0.183***
DYR3	0.183***	0.183***	0.183***	0.192***	0.183***	0.200***	0.050***
_cons	-0.139***	-0.139	-0.237***	-0.239***	-0.138***	-0.258***	-0.13***
Eff Score	0.88	1.00	0.9	0.91	0.88	0.9	0.82
*** p<.1; ** p<.05; *** p<.01**							

6. Conclusion

84. In this analysis, we used a 5-year balanced panel data (covering the period from 2011-12 to 2015-16) for Network Rail's ten routes to analyse the impact of different cost drivers on Network Rail's maintenance and renewal costs and to estimate the notional cost efficiency of different routes.

85. We obtained the data from Network Rail and cleaned it with Network Rail's support. We started the analysis by ensuring consistency and reliability of our data. We did this by conducting various checks including identifying potential outliers, missing or inconsistent data, adjusting for year-on-year fluctuations in renewal expenditure through a steady state adjustment, adjusting the cost data for inflation, etc. Any inconsistency was investigated and handled with the help of Network Rail. Then we discussed the trends in our data for the period under analysis.

86. Given the small size of our dataset as well as lack of data on some important cost drivers, we chose to base our conclusions on a simple but commonly used econometric approach, COLS, the results of which were checked against more advanced stochastic frontier analysis models. We conducted robustness checks by running models with different specifications of the cost variable (maintenance and

renewals cost together and then separately) as well as the traffic density variable (passenger train km and freight train km together and then separately). We also presented results with both steady state adjusted and non-steady state adjusted data. Overall, all the models give us comparable results. We also conducted a number of tests to check the validity of our model specification, and all of them showed that our model specification is valid.

87. Our preferred COLS model suggests that a 1% increase in track size and traffic density leads to an increase in maintenance and renewals costs by 0.95% and 0.8% respectively. We found evidence of the existence of economies of densities i.e. increasing traffic density increases the cost less than proportionally .This suggests that routes with more traffic density may have a cost advantage. Although the coefficient for track size (i.e. track km) was consistently below 1, which would suggest the existence of economies of scale, we could not conclude that the coefficient was statistically different from 1, suggesting that returns to scale may be constant. In addition, our analysis concludes that for a given length of track, it is cheaper to run it in multiple rather than a single track. Our results have also suggest that on average, routes spent more in the final year of CP4 than in any other year.

88. Our preferred COLS model produced a wide range of notional efficiency scores which fluctuate year on year. Everything else being equal, the difference between modelled frontier efficiency and route notional efficiency ranges from 8% to 22% but on average, it is modelled to be 16%.

89. This analysis has been useful in many ways. These include:

(i) It has helped us to improve our understanding of Network Rail's routes' characteristics (i.e. size, complexity, etc.) and spending behaviours (i.e. their financial performance, where routes are outliers, etc.)

(ii) It has laid the foundation for future analysis by identifying the weaknesses in our data and methodology that should be corrected in future analysis;

(iii) These results could be used to sense check results from other analysis within and outside ORR.

90. However, these results should be considered with caution as our analysis faced some constraints that may have significantly affected the robustness of our results. These include:

(i) The fact that we used a small data set;

(ii) Due to lack of data, we were not able to control for some variables that are commonly regarded as important drivers of costs (such as topography and route complexity, age of assets, etc.);

(iii) and the fact that we were unable to fully account for the lumpiness character of renewals expenditures, etc.

91. These constraints meant that we had to adopt a very simple COLS model as our main econometric technique. COLS models unrealistically assume that all the unexplained variation in cost is due to inefficiency. However, we know that in reality, the variation may also be due to different factors including measurement errors, omitted variables or other random occurrences. Although we catered for this by adjusting all our COLS efficiency scores upward by 25% (having regard to best practices in the regulatory industry), there is obviously some arbitrariness in choosing this uplift number and we consider that more sophisticated but direct approaches applied to a bigger dataset would have given us more credible results. We intend for this to be possible in our future analysis, as we will have longer data series.

92. Given these constraints, and for a more reliable future analysis in future control periods, we will work with Network Rail to look for ways to improve the quality of data available on all the cost drivers. We will support this by including the data needed for econometric top-down benchmarking in the ORR-Network Rail data protocol for CP6 and beyond.

93. As for the current analysis, these results should be used alongside other analysis to form an idea about Network Rail's efficiency. They are, in general terms, consistent with our view that there are opportunities for further efficiency savings.

Appendix B - Benchmarking of Maintenance Delivery Units

Econometric top-down benchmarking of Network Rail's maintenance delivery units (MDUs)

1. Introduction

1. As part of PR18, we have also undertaken the econometric top-down benchmarking of Network Rail's maintenance delivery units (MDUs). While this is a stand-alone analysis benchmarking maintenance costs for Network Rail's MDUs, it complements our top-down benchmarking of Network Rail's routes described in Appendix A. As explained in the econometric top-down benchmarking summary paper, these two strands of analysis are part of the PR18 efficient cost analysis whose ultimate aim is to set challenging but achievable efficiency savings targets for Network Rail during the next control period (CP6).

2. The main objective of this analysis is to compare the maintenance cost efficiency of Network Rail's MDUs. To achieve this, we use historical data to estimate a cost function and then produce efficiency scores for each MDU relative to the most efficient peer(s). For more details on this methodology and its assumptions, please see the analysis benchmarking Network Rail's routes in Appendix A.

3. In previous periodic reviews, we used international data to benchmark Network Rail's efficiency against similar infrastructure managers in Europe. In comparison, this intra-Network Rail benchmarking gives us insightful information about performance within Network Rail. MDUs (and routes) are easy to compare as they operate in broadly similar conditions.

4. MDUs are operating units within routes that are responsible for Network Rail's maintenance activity. At the time of our analysis, there were 37[10] MDUs. They accounted for nearly 70% of total network maintenance expenditure during the two years covered by this analysis i.e. 2014-15 and 2015-16. According to a Network Rail internal report (2012), MDUs:

   (i) inspect, service and maintain track, off track (line side vegetation, boundaries, drainage), signalling, electrical power and fixed plant assets;

   (ii) respond to unplanned infrastructure occurrences;

[10] The number and boundaries of MDUs change. For instance, Clapham, Eastleigh and Woking have recently become Wessex Inner and Wessex Outer. However, we keep them in our analysis as allocating the data to the two new MDUs may introduce further sources of errors in the data.

(iii) provide first line engineering support for work delivery;

(iv) plan and coordinate resource requirements for work delivery;

(v) manage and report maintenance activities; and

(vi) liaise with other stakeholders (internal and external) to minimise impact of maintenance activities.

5. In carrying out this analysis, we worked closely with Network Rail, who not only provided us with the necessary data but also worked constructively with us to cleanse it and correct errors and inconsistencies that we identified.

6. In this analysis, we applied the corrected ordinary least squares (COLS) methodology. This econometric top-down benchmarking approach is useful in identifying MDUs that perform most/less efficiently thereby informing the decision of where more scrutiny should be focused. However, econometric top-down benchmarking is not able to explain the reasons for the difference in efficiency among the MDUs. Moreover, as the scope and activities in each MDU are different, the high-level 'efficiency' scores produced by this approach may not fully reflect the exact level of efficiency.

7. This paper is organised as follows: after this introduction, section 2 discusses our data, trying to understand MDUs and their characteristics. Section 3 discusses our methodology. Section 4 presents our results while section 5 concludes.

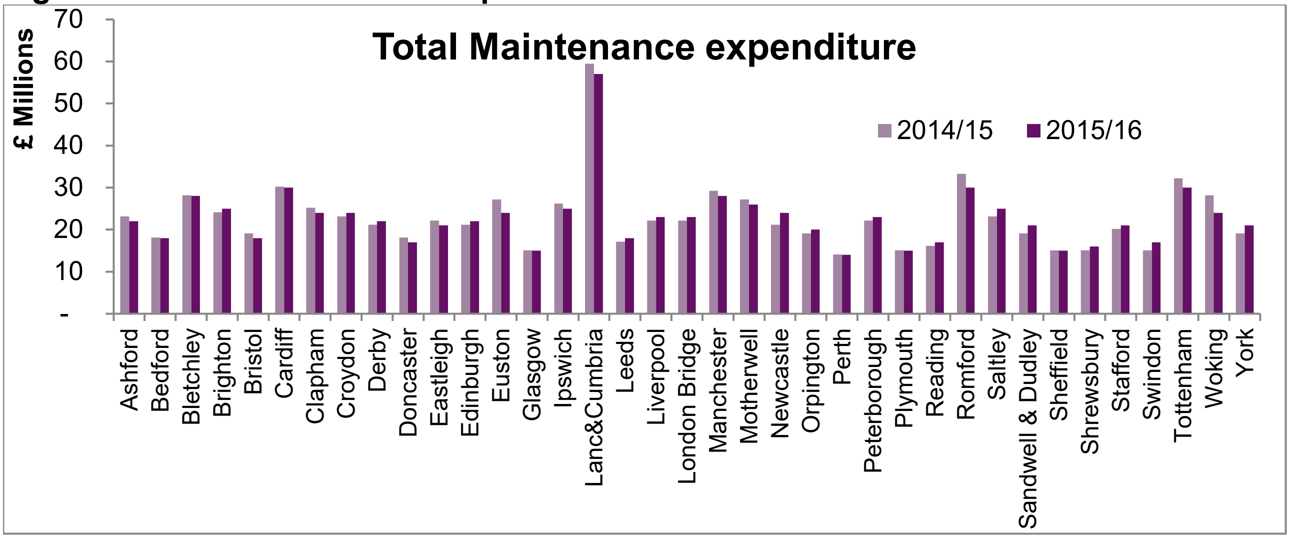
2. MDUs data and its sources

8. This section presents a few important variables (some of which we used in our econometric analysis) from Network Rail's MDU data. We discuss their sources and present trends in these variables to help understand differences in MDUs' behaviours and characteristics.

9. **Total maintenance expenditure**. MDUs accounted for 70% of total Network Rail maintenance expenditure in 2014-15 and 67.5% in 2015-16. So, over the two years covered by our analysis, only about 30% of Network Rail's maintenance expenditure was centrally managed. We obtained the data on total maintenance expenditure from Network Rail's regulatory accounts (2015-16) statement 8c. All the figures were in 2015-16 prices. Although the total amount spent on maintenance in MDUs was higher in 2015-16 than 2014-15 (from £1,198m to £1,248m), due to increased centrally managed spending, the total amount spent in MDUs was almost constant (£842m). Our analysis only looks at MDU spending. The data shows that the average MDU spent £23m per year. On average, Perth spent the least on maintenance (£14m); and Lancashire & Cumbria spent the most (£58m). The Lancashire & Cumbria data point looks like an outlier. As explained later in this document, in our analysis we run regressions with and without it to assess its impact. However, on

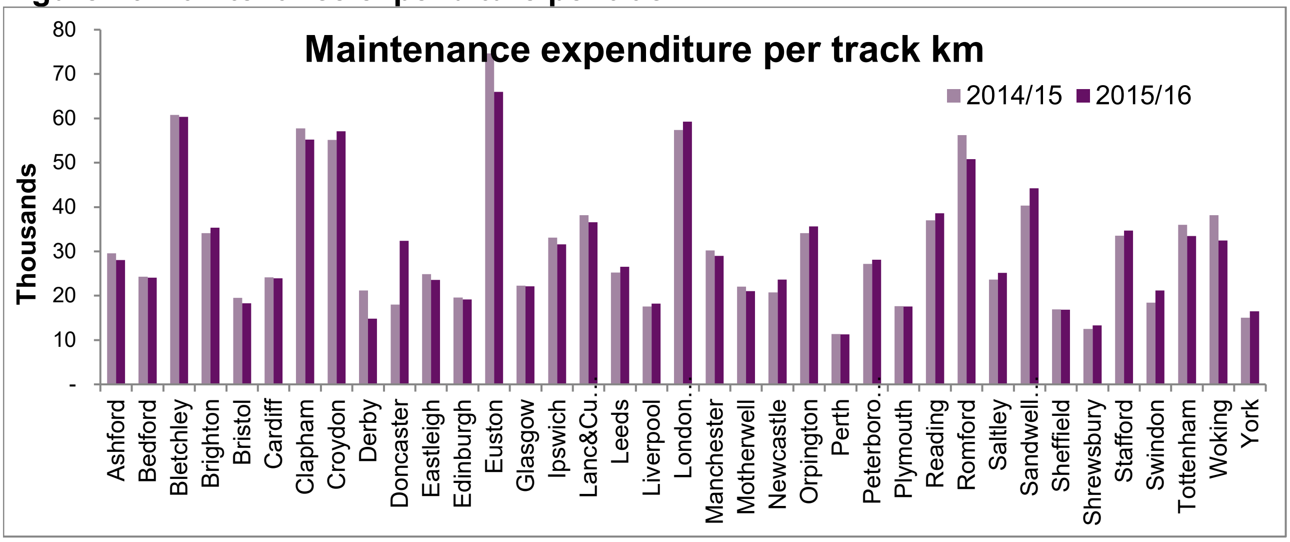
balance we decided to keep it in our final model because it is a genuine data point arising from the merger of two MDUs namely Preston and Carlisle. To reduce its impact we transformed the data by using the data in form of each variable's data point as a ratio of the average of all observations on that variable. We then used the data after transforming it into logarithms. This has greatly reduced any impact that any outlier could have on the results.

Figure 15 Total maintenance expenditure



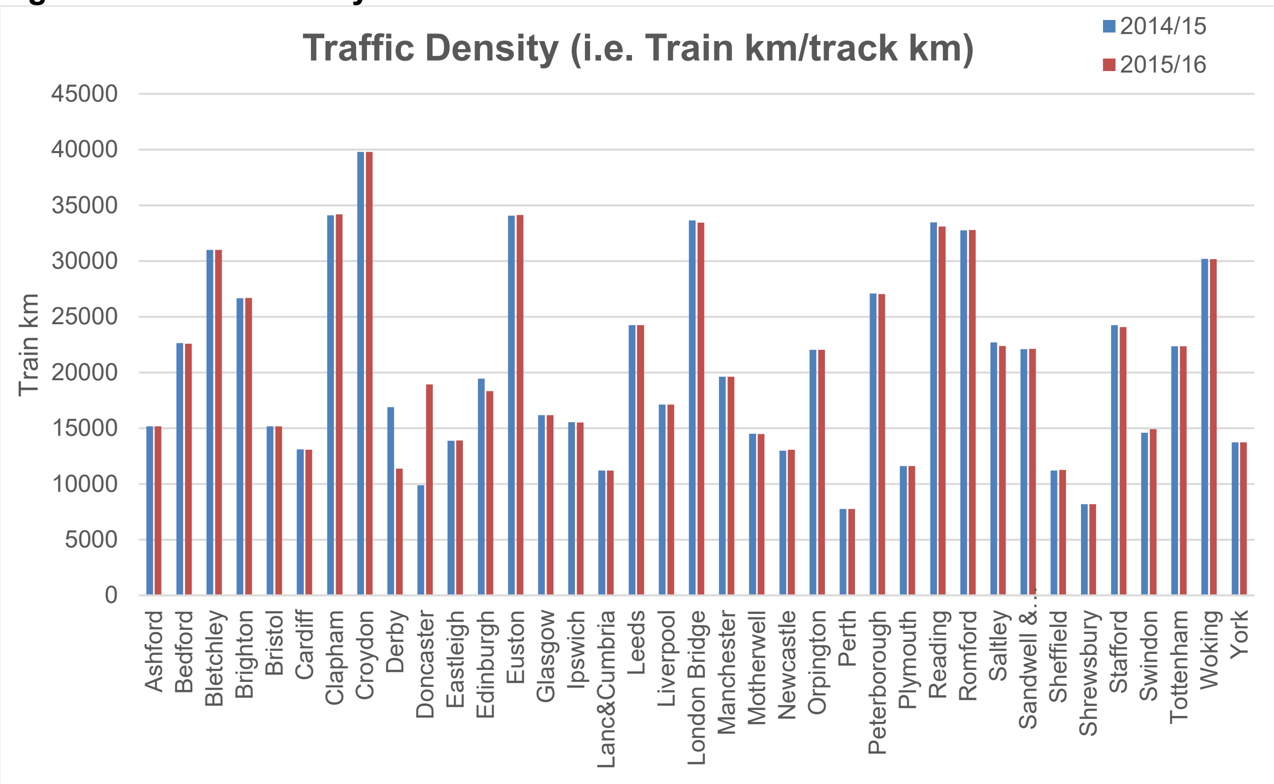
10. **Maintenance expenditure per track km**. We calculated maintenance expenditure per track km by dividing maintenance spending by the length of track in each MDU. An average MDU spent £31,000 to maintain 1 km of track. This measure should be considered carefully as the budget in question was not spent on track only. Bletchley, Clapham, Croydon, Euston, London Bridge and Romford spend above average while Perth has the lowest maintenance cost per track km at £11,000/ km

Figure 16 Maintenance expenditure per track km



11. **Traffic density**: We calculated the traffic density as train km (both passenger and freight) divided by track km in each MDU. Network Rail collected both track length and train km data through its assets management services. On average, traffic density was 18000 train km per track km. Croydon had the highest traffic density which partly explains its high maintenance cost per track km. Perth has the lowest density which may also explain its low maintenance cost per track km.

Figure 17 Traffic density



12. **Track km and other track characteristics**: Network Rail provided us with data on different network characteristics that they collect from their assets management services. According to this data, Lancashire & Cumbria and Derby are responsible for the longest tracks while Euston and London Bridge look after the shortest length of tracks. On average, the network in MDUs has 2 tracks with Euston, Peterborough and Reading having the highest number of average tracks (Track km/Route km) at 3.2. Average electrification was 50% (0% in Bristol, Cardiff, Perth, Plymouth and Swindon; and more than 95% in Clapham, Croydon, Euston, London Bridge, Orpington and Peterborough).

13. The network track is also classified according to the speed bands (miles per hour) i.e. the length of track in each of the following speed bands 0-35, 40-75, 80-105, and 110-125. For each MDU, we calculated the percentage (density) of the track in each band. Sheffield has the highest density of the lowest speed tracks (0-35) at 22%. Stratford has the highest density of the high-speed tracks (110-125 miles/hour) at 50%.

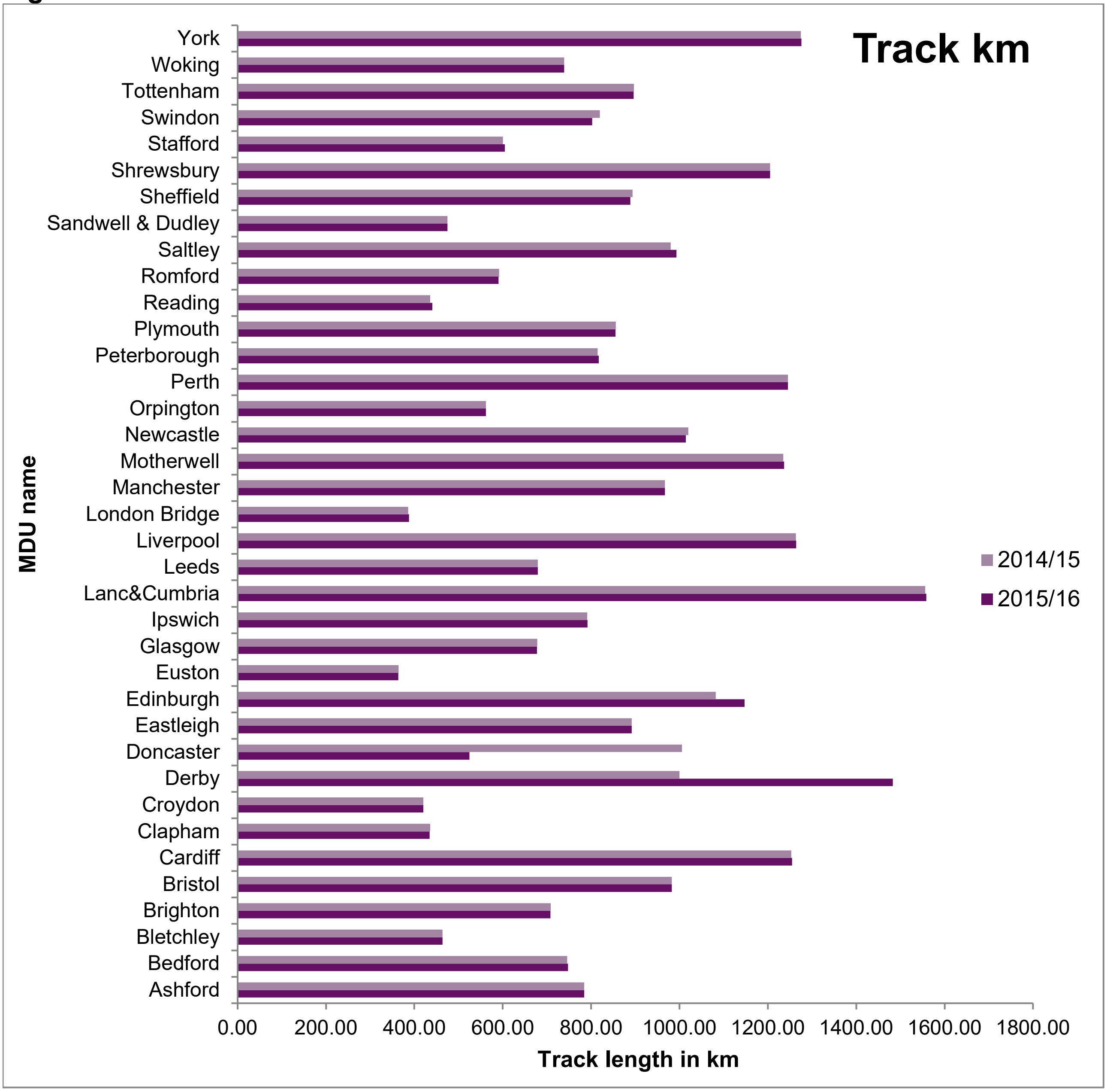
14. Tracks are further classified as primary, secondary and rural. Peterborough has the highest density of primary tracks (94%); Eastleigh has the highest secondary track density (77%); and Glasgow has almost 50% rural tracks.

15. Finally, tracks are classified based on criticality[11]. Bletchley has highest density of criticality 1 tracks (87%) while Peterborough has the highest density of criticality 2 tracks (60%). London Bridge has the highest density of criticality 3 tracks (70%), Tottenham the highest density of criticality 4 tracks (60%) and Shrewsbury the highest density of criticality 5 tracks (62%).

16. We found that there is high correlation between high-speed tracks, primary tracks and criticality 1 densities on the one hand, and between low-speed tracks, rural tracks and criticality 5 densities on the other. For example, Stafford and Bletchley have high densities of high-speed (110-125) tracks, primary tracks and criticality 1 tracks. However, this information should be interpreted carefully as a network section with high densities of rural, criticality 5 and low-speed tracks will not necessarily be cheaper to maintain. This is because there are other factors that influence costs (such as topography, age of the assets, weather, accessibility, etc.) which are not necessarily accounted for in those measures.

[11] Network Rail defines **Route criticality** as a "measure of the consequence of the infrastructure failing to perform its intended function, based on the historic cost of train delay per incident caused by the track asset". Using this measure, each strategic route section (SRS) of the network has been assigned a route criticality band from 1 to 5. The lower the number of the criticality band, the more a delay is likely to cost should infrastructure fail. The classification of each SRS into criticality bands is used in the development of Network Rail's asset policy as a first step to matching the timing and type of asset interventions.

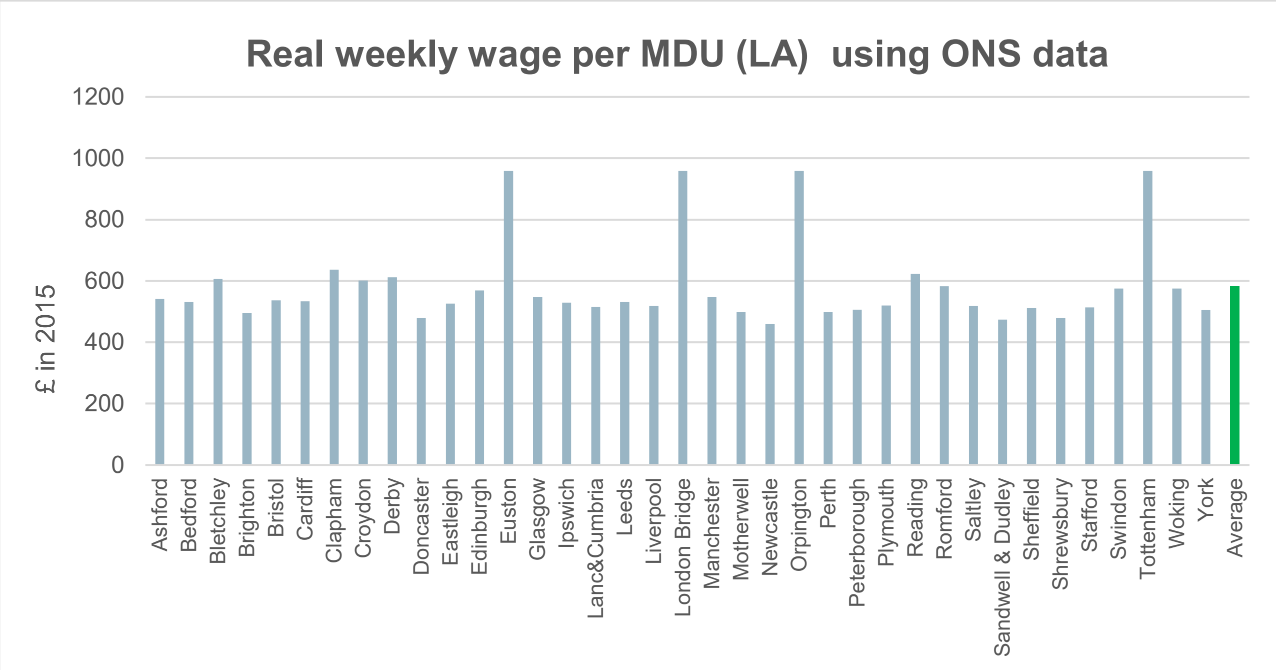
Figure 18 Track km



17. In our analysis, we also controlled for **labour input costs**. The Office of National statistics (ONS) publishes data on weekly earnings by local authority. We matched each of the 37 MDUs with local authorities in which they operate. This data is not railway specific and on average, Network Rail staff's weekly pay is higher than this. However, as we do not have data specific to Network Rail, we think that this data is good enough to gives us a broad idea of how expensive labour inputs (in general) are in each MDU's geographical area of operation. Similar approaches have been used in previously published academic research (e.g. see Wheat and Smith, 2008[i]). While our analysis could have benefited from controlling for prices of other inputs such as materials and machinery, we did not have the data to do so. However, we can assume that this is constant between MDUs and thus its effect could be accounted for within the constant term.

18. According to ONS, the data shows seasonally adjusted real average weekly earnings (AWE) per local authority. The data is in 2015 prices. The figure below visually compares labour input prices in the 37 MDUs geographical areas. The data suggests that labour inputs in Euston, London Bridge, Tottenham and Orpington should be more expensive while Newcastle, Shrewsbury and Doncaster should have the cheapest labour input.

Figure 19 Real weekly wage



3. Methodology

19. We analyse MDU maintenance cost efficiency using econometric top-down benchmarking. It involves establishing a cost function that explains differences in spending with a set of cost drivers. Comparators that most outperform this function form the 'efficient frontier'. By calculating the distance from this frontier, we can estimate the relative efficiency of the other comparators. The further they are from the frontier, the less efficient they are and therefore the greater is their scope for efficiency catch-up.

20. This technique compares each MDUs performance with the 'best in class' performer (s). Econometric top-down benchmarking also produces marginal costs for infrastructure use (i.e. elasticities from multiple cost drivers). However, the main objective of this analysis is to develop an approach for benchmarking MDUs' maintenance cost efficiency.

21. Econometric top-down benchmarking needs data of sufficient quantity and quality to produce reliable estimates. It may suffer from poor model specifications and functional form and does not allow for a qualitative understanding of the reasons for differences between companies. However, it may help to identify areas for further investigation and challenge. Our analysis and conclusions are mindful to these facts.

22. For more details on the econometric top-down benchmarking methodology and its assumptions, please refer to our analysis benchmarking Network Rail's routes in Appendix A.

23. Our methodology (choice of functional form, choice of variables to include in our analysis, etc.) was based on two strands of previous analysis: internal analysis by Network Rail (2010, 2011, and 2012), and published work by the Institute for Transport Studies-University of Leeds (Wheat and Smith, 2008[12]). Although those two studies had different objectives, they are both useful to our present analysis especially because they both focus on Network Rail's MDUs. In brief the two studies proceeded as follows:

24. Network Rail's internal analysis is the closest to ours in both its focus and objective. It applied ordinary least squares (OLS) regressions on 2010-11 data for 39 MDUs to estimate the relationship between maintenance costs and selected cost drivers. With total maintenance cost as their dependent variable, they controlled for the following cost drivers: structural factors (i.e. track length, level crossings, switches & crossings unit density, traffic volumes (train km/track km) and absolute track geometry (ATG)). According to Network Rail, their results "compare the best in class with the other delivery units and the difference between the actual and modelled budgets can be seen as potential for efficiency". The results were used to set the MDUs' budget with efficiency targets for the last 3 years of CP4 of around 16% i.e. £120m (Network Rail, 2012).

25. In their study, Wheat and Smith (2008)'s aim was to estimate the marginal cost of running more or less traffic on a fixed network in the UK. They applied ordinary least squares (OLS) to a cross section dataset of 53 MDUs in 2005-06. Their cost function relates total maintenance costs to output variables (train km/track km, tonne miles/track km, etc.), prices of inputs (labour, energy, etc.) and infrastructure capability/quality variables (route miles, number of switches and crossings, line speed, electrification, etc.). As the first step in top-down efficiency benchmarking is to estimate a cost function, this study is useful to us, especially in choosing the right independent variables.

[12] Wheat. P., and Smith, A.S.J. (2008), Assessing the Marginal Infrastructure Maintenance Wear and Tear Costs for Britain's Railway Network, Journal of Transport Economics and Policy, Vol. 42, No. 2 , pp. 189-224. Available at https://www.jstor.org/stable/20054045

26. In addition to those two studies, data availability was a major factor in our choice of methodology and which variables to include. In fact, Network Rail's collection of data at MDU level has not been consistent enough to provide us with all the variables we wish we could have included in our model (e.g. asset conditions and age). Therefore, given our small dataset (a two-year panel with 74 observations), the possibility of errors in measuring our variables, and the unavailability of data on some theoretically important cost drivers, our preferred model is a corrected ordinary least squares (COLS) model in a Cobb-Douglas double-log form as follows:

Ln(Maintenance Total Cost)= f(Lntrackkm + Lntraffic density_pax + Lntraffic density_fr + Lnwage + Electrified density+ Speed_ 40-75 density+ Ln average tracks+ Criticality_1 density) + random error

Where:

- ***Ln*** means 'natural logarithm';
- **track km** is the length of the track;
- **traffic density_pax** means passenger train km divided by track km;
- **traffic density_fr** means freight train km divided by track km;
- **average track** stands for track km divided by route km;
- **wage** stands for average real weekly earnings;
- **electrified density** is the proportion of track that is electrified;
- **speed_ 40-75 density** is the proportion of track with speed between 40-75 miles per hour; and
- **Criticality_1 density** is the proportion of track in criticality band

27. A Cobb-Douglas functional form is restrictive since it assumes constant cost elasticities. A translog cost function would be less restrictive, and theoretically more useful, as it incorporates additional second order interaction terms. But, in our view, a Cobb-Douglas functional form is more suitable to our data because of its size and low levels of variability.

28. Our MDU analysis applies the same methodology as our route level analysis. However, in addition to track size, traffic density and average track, here we also control for electrification density, speed 40-75 density and criticality_1 density. We expect those variables to have a positive relationship with the costs of maintaining the network within MDUs. Our expectation is that a highly electrified network as well as that with high density of criticality_1 track is more costly to maintain.

29. The main weakness in using a COLS model in analysing efficiency is that it assumes that any deviation from the efficiency frontier is only explained by inefficiency. This is a very strong (and perhaps unrealistic) assumption especially when dealing with a

data like ours i.e. a small dataset where the likelihood of measurement errors is high. Therefore, following the practice by other regulators (such as Ofwat and Ofgem as well as ORR's previously published analysis), we decided to uplift all the efficiency scores by 25%.[13]

30. During the analysis, we conducted different statistical tests to check the validity of our model specification. These include tests of skewness, multicollinearity, and heteroscedasticity. None of the tests[14] suggested that our model specification is invalid.

4. Results

31. This section presents the results from our model. As mentioned earlier, our analysis faced some significant data quality and quantity constraints. This means that these results have to be considered with caution, as we do not think they are robust enough to be used alone to draw strong conclusions about individual MDUs' efficiency. Therefore, while we expect our future analysis to produce more robust results, we recommend that in PR18, these results are used only to support and sense check results from other analyses.

32. Our model's results are in Table 13 below. All variables in the model are statistically significant and coefficients have the expected signs. The model suggests the existence of economies of scale and density. This means that increasing network size and traffic density increases the cost less than proportionally: MDUs with longer networks and those with more densely used networks have a cost advantage. Specifically, our model suggests that increasing track length by 1% increases the maintenance cost by 0.54%. Increasing passenger and freight traffic density by 1% increases maintenance cost by 0.57% and 0.13% respectively. Track length and traffic density are both statistically significant at 1%. Electrification and wage are positively related to maintenance cost and are statistically significant at 10%. As expected, increasing the proportion of the track that is in criticality band 1 increases maintenance cost and the variable is statistically significant at 1%. Similarly, our model suggests that it is cheaper to run the network in multiple tracks rather than in single track.

[13] On the benefits of using COLS as compared to other complicated models such as Stochastic Frontier analysis, as well as rationale of the adjustments made on the efficiency scores obtained from it, please refer to the discussion in our report on benchmarking Network Rail's routes

[14] Available on demand

Table 13 OLS results and COLS notional efficiency scores for our main model

Maintenance total	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
LnTrackkm	0.540	0.120	4.5	0.000	0.300	0.780
Electified_density	0.120	0.047	2.54	0.013	0.026	0.214
LnTraffic_density_pax	0.566	0.110	5.16	0.000	0.347	0.785
LnTraffic_density_fr	0.132	0.041	3.2	0.002	0.050	0.215
Speed_40_75_km	0.264	0.071	3.74	0.000	0.123	0.406
LnAvg_tracks	-0.372	0.209	-1.78	0.079	-0.789	0.045
Critical_1_km	0.209	0.046	4.52	0.000	0.117	0.301
LnWage_ons	0.202	0.119	1.7	0.095	-0.036	0.440
_cons	-0.051	0.056	-0.9	0.369	-0.162	0.061
	Obs.	Mean	Std. Dev.	Min	Max	
25% upward adj. eff_score	74	0.84	0.34	0.63	1.00	

33. Our COLS model produced a wide range of notional efficiency scores which fluctuate year on year. Everything else being equal, the difference between modelled frontier efficiency and notional efficiency scores for particular MDUs ranges from 3% to 37% but on average, it is modelled to be 16%. We present this in table 14 (which also maps MDUs with the operating routes to which they belong) and visually in Figure 20 below.

34. The main shortcoming with using COLS to analyse efficiency is that it assumes that any deviation from the frontier is only explained as inefficiency. However, this is a very strong assumption, especially when analysing a small dataset with high likelihood of measurement errors like in the present case. We had the option to adopt more complicated approaches such as stochastic frontier analysis (SFA), which use more realistic assumptions. But we chose to use COLS as, in our view, its simplicity means it is the most appropriate to analyse the kind of data at our disposal. However, to reduce the impact of the COLS strong assumption, we adjusted each efficiency score by 25% assuming that 25% of the deviation from the frontier results from random noise. This practice is widely used by regulators and we used it in our previous benchmarking[15] work in PR13.

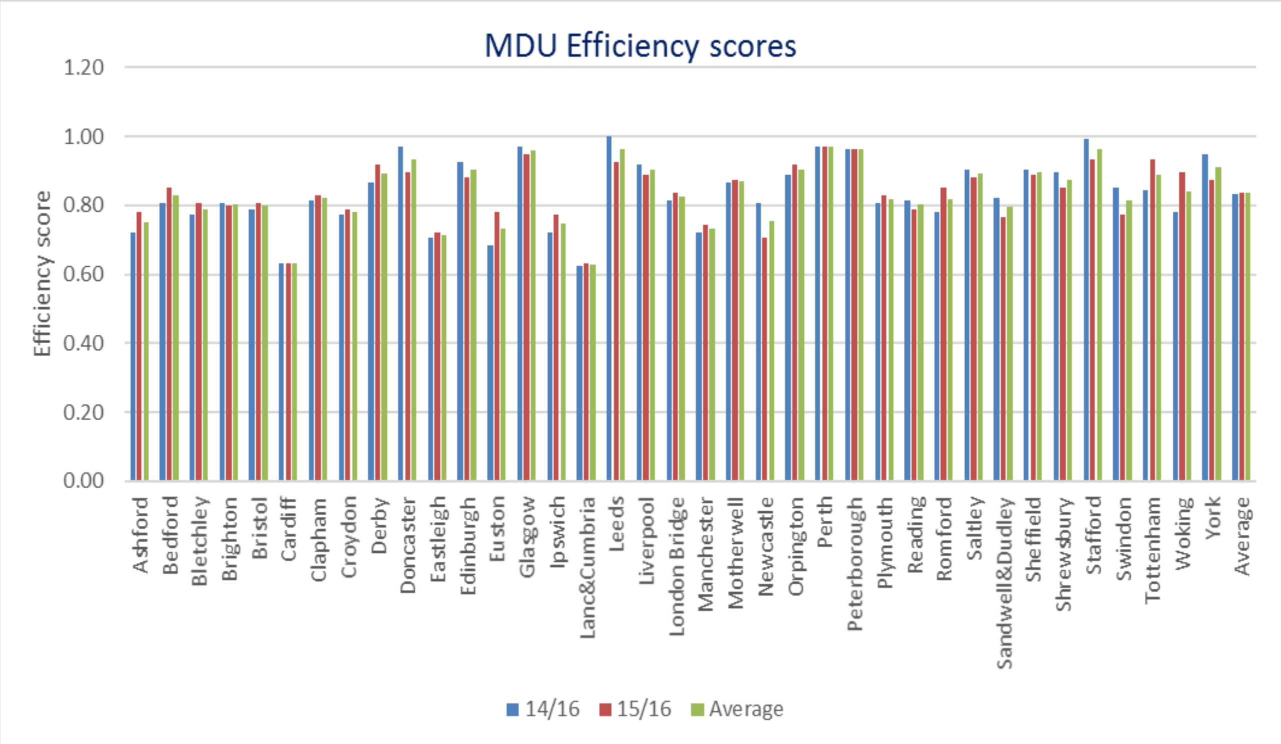
[15] For more details, please see the Benchmarking of routes in appendix A

Table 14 Modelled individual MDUs notional efficiency scores

Route	MDU	2014-15	2015-16	Average
Anglia	Ipswich	0.72	0.78	0.75
	Romford	0.78	0.85	0.82
	Tottenham	0.84	0.93	0.89
EM	Bedford	0.81	0.85	0.83
	Derby	0.87	0.92	0.89
LNE	Doncaster	0.97	0.90	0.93
	Leeds	1.00	0.93	0.96
	Newcastle	0.81	0.71	0.76
	Peterborough	0.96	0.96	0.96
	Sheffield	0.90	0.89	0.90
	York	0.95	0.87	0.91
LNW	Bletchley	0.78	0.81	0.79
	Euston	0.69	0.78	0.73
	Lanc & Cumbria	0.63	0.63	0.63
	Liverpool	0.92	0.89	0.90
	Manchester	0.72	0.75	0.73
	Saltley	0.90	0.88	0.89
	Sandwell & Dudley	0.82	0.77	0.79
	Stafford	0.99	0.93	0.96
Scotland	Edinburgh	0.93	0.88	0.90
	Glasgow	0.97	0.95	0.96
	Motherwell	0.87	0.87	0.87
	Perth	0.97	0.97	0.97
Sussex	Croydon	0.78	0.79	0.78
	Brighton	0.81	0.80	0.80
Kent	Ashford	0.72	0.78	0.75
	London Bridge	0.81	0.84	0.82
	Orpington	0.89	0.92	0.90

Route	MDU	2014-15	2015-16	Average
Wales	Cardiff	0.63	0.63	0.63
	Shrewsbury	0.90	0.85	0.87
Wessex	Clapham	0.81	0.83	0.82
	Eastleigh	0.71	0.72	0.72
	Woking	0.78	0.90	0.84
Western	Bristol	0.79	0.81	0.80
	Plymouth	0.81	0.83	0.82
	Reading	0.81	0.79	0.80
	Swindon	0.85	0.78	0.81
	Average	0.83	0.84	0.84

Figure 20 MDUs' notional efficiency scores



5. Conclusion

35. About 70% of Network Rail's maintenance budget is spent in maintenance delivery units (MDUs) that are responsible for maintenance activities within Network Rail's routes. In this paper, we applied econometric techniques to a two-year data (2014-15 and 2015-16) of the sort that could be used to benchmark Network Rail's 37 MDUs (that existed at the time of analysis) in terms of their spending efficiency as compared to the modelled frontier efficiency. The analysis drew from two strands of previously conducted analyses by Network Rail (2010, 2011, and 2012) and by the Institute for Transport Studies-University of Leeds (Wheat and Smith, 2008).

36. This is the first time that ORR has conducted such an analysis at such a geographically disaggregated level. This analysis constitutes an important step towards increased regulation of Network Rail at a route level. It also constitutes a basis for future analysis as it has identified data issues that ORR and Network Rail will need to handle to ensure better quality data is available for future analysis.

37. Given the small size of the dataset as well as our inability to obtain data on some potential cost drivers (and therefore our inability to control for them in our regression), we adopted a simple but widely used methodology i.e. the corrected ordinary least squares (COLS) methodology to estimate the relationship between total maintenance cost and its drivers, and then we estimated the efficiency scores for each MDU as compared to the modelled frontier efficiency

38. Our model suggests economies of scale and economies of densities. This means that increasing network size and traffic density increases the cost less than proportionally. If the results of this model are accurate, then MDUs that maintain larger and more densely used networks have a cost advantage. Our model also suggests that it is cheaper to maintain infrastructure with multiple, rather than single, tracks.

39. Our model produced a wide range of notional efficiency scores which fluctuate year on year. Everything else being equal, our model estimates that the difference between modelled frontier efficiency and MDUs' notional efficiency scores ranges from 3% to 37% but on average, it is modelled to be 16%. These results are comparable to the results of our route data analysis.

40. We conducted various statistical tests to check the validity of our model specification. None of these results suggested that our model was invalid. We believe that this model is robust from an econometric perspective.

41. However, this analysis faced some significant data quantity and quality constraints. This means that our modelled efficiency frontier may not represent the exact MDUs cost structure. Consequently, we do not think these results are robust enough to be

used alone to decide on MDUs' (or Network Rail's) efficiency target for CP6. However, they are robust enough to be used alongside and to sense-check evidence from other analyses (such as bottom-up benchmarking) which inform that decision.

42. To ensure that better quality data is available for future analysis we recommend that the data we need for econometric top-down benchmarking of MDUs be included in the ORR-Network Rail data protocol for CP6 and beyond.